DOCUMENT RESUME

ED 231 402 IR 050 387

AUTHOR	Tolle, John E.; And Others
TITLE	Current Utilization of Online Catalogs: Transaction Log Analysis. Volume I of Three Volumes. Final Report.
INSTITUTION	OCLC Online Computer Library Center, Inc., Dublin, Ohio.
SPONS AGENCY	Council on Library Resources, Inc., Washington, D.C.
REPORT NO	OCLC/OPR/RR-83/2
PUB DATE	31 Mar 83
NOTE	127p.; For related documents, see IR 050 387-389.
AVAILABLE FROM	OCLC, Dept. Box ONB, Columbus, OH 43265 ($6.50 per copy prepaid).
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	Data Collection; Library Automation; *Library Catalogs; Library Research; *Online Systems; *Research Methodology; *Search Strategies; Use Studies
IDENTIFIERS	*Online Catalogs; *Transaction Log Analysis

ABSTRACT

As a continuation of research into online public access catalog systems (OPAC's), this study focused on the actual use of OPAC's and on refinement of the methodologies used in carrying out transaction analysis studies. Current patron use patterns were obtained via machine-readable catalogs from four online public catalog systems: The Library of Congress, Syracuse University, Dallas Public Library, and The Ohio State University. Data were also collected in the searching of 76 OCLC public terminals at 53 additional libraries. Such data included terminal identification, user commands and codes, system commands and codes, the search key, session beginning and ending, user and system response times, search text, and dates. Results indicate that accurate information on the amount of time spent at the terminals is unavailable except at The Ohio State University; time spent at terminals is higher than time spent at card catalogs at the same institution; errors frequently occur in sequence; and transaction log analysis is an effective tool for studying user behavior. Utilizing the OPAC transaction tapes, statistical measures have established the relative use of commands within the systems and the sequences of actions to determine user patterns. This report includes 71 illustrations and 46 tables. (Author/LMM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED231402

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Report Number: OCLC/OPR/RR-83/2
1983 March 31

Final Report to the Council on Library Resources

VOLUME I

of Three Volumes

CURRENT UTILIZATION OF
ONLINE CATALOGS:
TRANSACTION LOG ANALYSIS

by

John E. Tolle, Ph.D.
Project Manager

with
Research Assistance
from
Robert Allison, Sehchang Hah, and Walter Johnson

IR050387

OCLC Online Computer Library Center, Inc.
Office of Research
6565 Frantz Road
Dublin, OH 43017-0702

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Lois Yoakum

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

All rights reserved; however, individuals may quote from or reproduce portions of the text for noncommercial, educational, or private study or research. Full credit must be given to both the author(s) and the Office of Research, OCLC, Dublin, Ohio.

RESEARCH REPORT SERIES

Through the Research Report Series, OCLC research project results are made available to the public. OCLC Research Reports are published and distributed by OCLC. In addition, the reports are deposited with ERIC, which makes them available in microfiche as well as paper copy. Until Research Reports are available from ERIC (usually six to nine months after publication), OCLC will supply copies to persons and institutions requesting them. Due to the increasing demand for these reports and increasing costs to produce them, OCLC charges requestors a nominal fee. The OCLC price or ERIC number is noted below for each report. Currently, orders for OCLC Research Reports must be prepaid; send orders and checks to OCLC, Dept. Box ONB, Columbus, OH 43265.

Subject Heading Patterns in OCLC Monographic Records
by E. O'Neill and R. Aluri
Report Number: OCLC/RDD/RR-79/1; ERIC ED 183167

An Overview of a Proposed Monitoring Facility for the Large-scale, Network-based OCLC Online System
by W. Dominick, D. Penniman, and J. Rush
Report Number: OCLC/OPR/RR-80/1; ERIC ED 186042

Analytical Review of Catalog Use Studies
by K. Markey
Report Number: OCLC/OPR/RR-80/2; ERIC ED 186041

A Method for Correcting Typographical Errors in Subject Headings in OCLC Records
by E. O'Neill and R. Aluri
Report Number: OCLC/OPR/RR-80/3; ERIC ED 215680

Field, Subfield, and Indicator Statistics in OCLC Bibliographic Records
by T. B. Hickey
Report Number: OCLC/OPR/RR-81/1; ERIC ED 202484

Development of a Probabilistic Author Search and Matching Technique for Retrieval and Creation of Bibliographic Records
by T. B. Hickey
Report Number: OCLC/OPR/RR-81/2; ERIC ED 206286

Survey of Online Searching Instruction in Schools of Library and Information Science
by C. H. Fenichel and S. P. Harter
Report Number: OCLC/DD/RR-81/3; ERIC ED 210038

The American Public Library
by D. Sager
Report Number: OCLC/OPR/RR-82/1; ERIC ED 221222

The Process of Subject Searching in the Library Catalog: Final Report of the Subject Access Research Project
by K. Markey
Report Number: OCLC/OPR/RR-83/1
OCLC Price: $7.00; 152 pages

Current Utilization of Online Catalogs: Transaction Log Analysis;
Final Report to the Council on Library Resources, Volume I
by J. E. Tolle
Report Number: OCLC/OPR/RR-83/2
OCLC Price: $6.50; 114 pages

Online Catalog Use: Results of Surveys and Focus Group Interviews in Several Libraries;
Final Report to the Council on Library Resources, Volume II
by K. Markey
Report Number: OCLC/OPR/RR-83/3
OCLC Price: $12.50; 284 pages

A Comprehensive Study of Online Public Access Catalogs:
An Overview and Application of Findings;
Final Report to the Council on Library Resources, Volume III
by N. K. Kaske and N. P. Sanders
Report Number: OCLC/OPR/RR-83/4
OCLC Price: $5.50; 102 pages

OCLC LIBRARY, INFORMATION, AND COMPUTER SCIENCE SERIES

The OCLC Library, Information, and Computer Science Series of monographs is another vehicle through which the OCLC Office of Research communicates the results of significant research investigations and discusses issues of major importance. The first monograph was published in June 1982; the second will be available in spring 1983. More titles are planned for the near future. To obtain an order form or additional information about this series, contact OCLC, Documentation Department, 6565 Frantz Road, Dublin, OH 43017-0702; send orders and checks for monographs to the address noted above for OCLC Research Reports.

Online Public Access Catalogs: The User Interface
by C. R. Hildreth
Product Number: 801
Price: $18.00; 280 pages, paperbound

Public Library Administrators' Planning Guide to Automation
by D. J. Sager
Forthcoming

NOTE TO READER

Three volumes comprise the Final Report to the Council on Library Resources, Inc., prepared by the OCLC Office of Research.

Volume I.	Current Utilization of Online Catalogs: Transaction Log Analysis, by John E. Tolle.
Volume II.	Online Catalog Use: Results of Surveys and Focus Group Interviews in Several Libraries, by Karen Markey.
Volume III.	A Comprehensive Study of Online Public Access Catalogs: An Overview and Application of Findings, by Neal K. Kaske and Nancy P. Sanders.

In addition to these volumes, the reader may wish to consult the final reports prepared by the organizations which were also funded by the Council on Library Resources, Inc. in this online public access catalog study. Following is a list of these reports:

Research Libraries Group, Inc. Public online catalogs and research libraries: final report to the Council on Library Resources. Stanford, CA: Research Libraries Group, Inc., 1982 September.

Matthews, Joseph K. A study of six online public access catalogs: a final report submitted to the Council on Library Resources, Inc. Grass Valley, CA: J. Matthews & Associates, Inc.; 1982 November.

University of California, Division of Library Automation and Library Research and Analysis Group. Users look at online catalogs: results of a national survey of users and non-users of online public access catalogs: final report to the Council on Library Resources. Berkeley, CA: University of California; 1982 November 16.

Anderson, Rosemary; Reich, Victoria A.; Wagner, Pamela Roper; Zich, Robert. Library of Congress online public catalog users survey: a report to the Council on Library Resources. Washington, DC: Library of Congress, Office of Planning and Development; 1982 October 29.

ACKNOWLEDGMENTS

The following individuals contributed directly to the research effort through numerous project meetings, data reduction, programming efforts, general brainstorming, and project discussions. Charles Hildreth, OCLC Research Scientist, reviewed the attempts to define the taxonomy of state selections for the systems, helped define the final taxonomies, and put together the OCLC system user statistics. Walter Johnson, Bob Allison, Sehchang Hah, and Dave Hsaio, all Research Assistants in the Office of Research, were responsible for data reduction of the transaction tapes. Walter Johnson concentrated principally on Syracuse University, Bob Allison on the Library of Congress, and Sehchang Han on Dallas Public Library. Bob Allison and Christine Borgman of OCLC both worked on the analysis of The Ohio State University data. We discussed our activities with one another extensively, since the dynamic nature of the project required current feedback and occasional adjustment to our efforts. The analytic programs for chaining were written by Howard Turtle, Senior Research Scientist, and modified by Trong Do, Programmer Analyst, both of the OCLC Office of Research.

In addition to the above individuals directly involved in the research, The Council on Library Resources, Inc. is acknowledged for its financial support, in particular Project Monitor C. Lee Jones. Acknowledgement is also due to OCLC and Neal K. Kaske, Director of OCLC's Office of Research, who gave their support to the project. David Penniman, Vice President of Planning and Research, OCLC, encouraged the work on transaction analysis and contributed directly through his earlier efforts in transaction analysis. Thanks go also to Karen Markey, Research Scientist, for her discussions and comments throughout the project.

The assistance from the participating libraries is also gratefully acknowledged. The support of Pam Bacher, the Library of Congress; John C. Wyman, Manager of Systems Development, Syracuse University Libraries; Antoinette L. Johnson, Librarian for Systems Development, Dallas Public Library; William J. Studer, Director, The Ohio State University Libraries and Susan Logan, Assistant Professor, The Ohio State University Libraries and their staffs was critical to the success of the project. Special thanks are offered to Professor Pauline Cochrane of Syracuse University for her discussions and encouragement.

Appreciation is expressed to the OCLC staff for their extensive assistance without which this report could not have been completed. Cindy Heffelfinger did the typing with assistance from Marifay Makssour, Mary Taylor and Lorelei Yoder. Numerous individuals provided editorial assistance including Sig Dierk, Debbie Harman, Cray Little, Jan Lucas and Leslie Pearse. Richard Skopin and Rick Limes of the graphics area put in extended effort to complete the graphics for this report.

NOTES ABOUT THE AUTHOR

John E. Tolle is a Research Scientist in OCLC's Office of Research. He recently completed a study for the National Science Foundation on Terminal Requirements for Online Catalogs and is preparing a monograph on this topic for the OCLC Library Information and Computer Science Series. In addition to his active research in transaction log analysis of online catalogs, he is involved in the Improved Access Techniques and Microcomputer Applications programs. Dr. Tolle has conducted extensive investigations into the areas of systems engineering, analysis, and operations. Prior to joining OCLC, he taught and conducted research at Carnegie-Mellon University and Battelle Columbus Laboratories. He presently serves as adjunct Associate Professor at the Ohio State University. Dr. Tolle earned his B.Sc. and M.Sc. in Mathematics, and his Ph.D. in Systems Engineering.

ABSTRACT

In November 1981, the OCLC Online Computer Library Center, Inc., Office of Research proposed to the Council on Library Resources, Inc., to continue and expand its research into Online Public Access Catalog systems (OPACs).

The goals were to obtain a better understanding of the actual use of online catalogs and to refine the methodologies used in carrying out any transition analysis study. The objective of transaction log analysis is to determine to what extent the current OPAC system features are used and the corresponding patterns of use. Current patron use patterns have been obtained via transaction log tapes, i.e., machine-readable logs from several online public catalog systems: the Library of Congress, Syracuse University, Dallas Public Library, and The Ohio State University. Utilizing the transaction tapes from the online public catalogs, statistical measures have established relative use of commands within systems and the sequences of actions to establish user patterns.

CONTENTS

NOTE TO READER, v

ACKNOWLEDGMENTS, vi

NOTES ABOUT THE AUTHOR, vii

ABSTRACT, viii

LIST OF ILLUSTRATIONS, xi

LIST OF TABLES, xiii

SUMMARY OF THE REPORT, 1

1.0 INTRODUCTION, 9

- 1.1 Project Objective, 9
- 1.2 Project Scope, 9

2.0 DATA COLLECTION, 11

- 2.1 Library of Congress, 13
- 2.2 Syracuse University, 13
- 2.3 Dallas Public Library, 13
- 2.4 The Ohio State University, 13
- 2.5 OCLC Public Terminals, 14

3.0 METHODOLOGY, 15

- 3.1 Problem Structure, 15
- 3.2 Transaction Tape Reduction and Parsing: Session Determination, 17
- 3.3 Taxonomy of State Selection, System Search Commands, and State Mappings, 18
- 3.4 Analysis Methodologies, 20
  - 3.4.1 Markov Analysis of State Chains, 20
  - 3.4.2 Transition Probability Matrices (TPMs), 21
  - 3.4.3 Kolmogorov-Smirnov Test, 22

4.0 RESULTS, 23

- 4.1 Proposed Hypotheses/Questions, 23
- 4.2 Library of Congress, 24
  - 4.2.1 Transition Probability Matrices, State Diagrams, and Transaction Statistics, 29
- 4.3 Syracuse University, 39
  - 4.3.1 Transition Probability Matrices, State Diagrams, and Transaction Statistics, 46
- 4.4 Dallas Public Library, 64

    - 4.4.1 Transition Probability Matrices, State Diagrams, and Transaction Statistics, 67
  - 4.5 The Ohio State University, 77
    - 4.5.1 Transition Probability Matrices, State Diagrams, and Transaction Statistics, 77
  - 4.6 OCLC Public Terminals, 99

5.0 CONCLUSIONS, OBSERVATIONS, AND RECOMMENDATIONS FOR FURTHER STUDY, 105

  - 5.1 Questions and Answers, 105
  - 5.2 Other Observations and Conclusions, 109
  - 5.3 Transaction Analysis Goals, 112
  - 5.4 Future Efforts, 112

LIST OF ILLUSTRATIONS

Figure	Title
1	Flow Diagram for Transaction Analysis Methodology, 16
2	Transaction Log Record Format, 17
3	Primary Mapping for State Selection for General Use, 19
4-13	State Transition Diagrams - Library of Congress, 34-38
14-24	State Transition Diagrams - Syracuse (SULIRS) University, 58-63
25-34	State Transition Diagrams - Dallas Public Library, 72-76
35-66	State Transition Diagrams - Ohio State University, 83-98
67	OCLC Public Terminals in Use, 101
68	OCLC Public Terminal Transactions, 101
69	Frequency of OCLC Searches by Type (May-Oct. 1982), 102
70	OCLC Public Terminals: Author and Title Searches (May-Oct. 1982), 102
71	OCLC Public Terminals: Standard Number Searches (May-Oct. 1982), 103

LIST OF TABLES

Table	Title
1	Summary of Empirical Data Log Elements, 12
2	Primary Mapping SCORPIO - LC States/Commands, 25
3	User Commands Available in SCORPIO - LC, 27
4	Sample Search in SCORPIO - LC, 28
5	State Frequency Distribution - Primary Mapping 0th Order - Context Validated - LC, 30
6-12	Transition Probability Matrices 1st through 4th Orders - LC, 30-33
13	SULIRS - STATES/COMMANDS, 44
14-30	Transition Frequencies and State Transition Matrices - SULIRS (Syracuse), 47-55
31	Frequencies of Certain Ranges of Hits Given Specific Search Types - SULIRS, 55
32	Percent of Range of Hits Given Specific Search Types - SULIRS, 56
33	Frequency of Repetitions of Specific Search Commands - SULIRS, 57
34	Frequency of Time-Gaps for N Codes - Dallas, 68
35	0th Order Matrix - Primary Mapping - Dallas, 68
36	First Order Transition Probability Matrix Primary Mapping - Successor - Dallas, 69
37	Second Order Transition Probability Matrix Primary Mapping - Successor - Dallas, 69
38	Summary of Search Types Within Sessions - Dallas, 70
39	Frequency Distribution for Search Types - Dallas, 71
40-43	State Transition Matrices - Compatible 1st through 4th Orders - OSU, 78-79
44-46	State Transition Matrices - Fifth Alternate 1st through 3rd Orders - OSU, 80-82

Report Number: OCLC/OPR/RR-83/2
1983 March 31

SUMMARY OF THE REPORT

A. OBJECTIVE AND GOALS

Part I of this study is concerned with the analysis of the transaction log records of online catalogs. The objective of this part of the online public access catalog study was to discover the extent to which current online catalog features are used and patterns of that use. While examining the actual catalog use, at least two goals became obvious: to obtain a better understanding of the use of online catalogs and to refine the methodologies used in carrying out any transaction analysis study.

The first goal, to obtain a better understanding of the actual online catalog use, may seem to be too large or vague, as the objective is contained within it. However, it is important to keep in mind that a valuable asset of the research is that it generates questions that were not previously considered. As such, it contributes to an understanding of the catalog use. This study has achieved this in the sense that considerable understanding of catalog use has been obtained. There is of course more to do and investigate, but we are much closer to understanding how the online catalog is used.

The second goal, to refine the methodologies, was certainly achieved. The use of transaction analysis--via development of transition probability matrices, state definitions, state diagrams, and Markov analysis--to create chains of user patterns certainly offers promise. This transaction analysis methodology is still in the early stages of refinement, however. Further refinement will occur as more systems are examined and more data is analyzed.

The scope of the transaction analysis included collecting data from four different online catalog systems. Machine-readable log tapes, which contained the user transactions in various formats and codes, were received from the following institutions: Library of Congress, Syracuse University, Dallas Public Library, and The Ohio State University. In addition, data was collected on the searching of 76 OCLC public terminals at 53 different libraries (as defined by unique OCLC location symbols). By institution type, these 53 libraries included 5 public libraries, 5 federal or state libraries, and 43 academic libraries (including several medical, health, and law libraries).

B. DATA COLLECTION

For the other online catalog systems, we obtained transaction log tapes covering the following periods:

Library of Congress - March to June 1982
Syracuse University - April 12 to April 28, 1982
Dallas Public Library - August 9 to August 21, 1982
The Ohio State University - January to December 1981, March to May 1982, and November 1982

Originally we requested two months of data from each library, but the actual data received ranged from two weeks to one full year. Using this data, samples were selected for analysis. For Syracuse University and Dallas Public Library, data was collected for two weeks, and included an examination of the representativeness of subsets of data. For Syracuse, some of the days were eliminated due to some mislabeled tapes and time consideration. For the Library of Congress, public use data was analyzed for the months of March 1982 and June 1982. April 1982 contained staff use data, and May 1982 had a mixture of both staff and public use data. These months were not analyzed, since we were interested solely in patron usage. Spring quarter 1981 was used for transaction analysis at The Ohio State University because session boundaries were defined by observers stationed at the catalog.

Each online catalog system was different; therefore, the resulting transaction log tapes varied in content and format. These differences are quite extensive, and the data reduction, coding, and reformatting represented a major task within the project. It was estimated, after the fact, that in conducting any transaction log analysis, well over 50% of the effort in labor and computer time (including programming, tape reduction, and reformatting) can be attributed to this phase.

Specific data elements collected (where available) included the following: terminal identification, user commands and codes, system commands and codes, the search key, session beginning, session ending, times of user and system responses, search text, and dates.

The above list represents an ideal rather than the reality, since no OPAC system studied had all of these elements as part of the information captured on the system logs and copied to tapes for analysis. In order to appreciate the differences among the systems studied, consider that Ohio State, Syracuse, and Dallas Public did not have any session determinates, i.e., session length (when a user begins a session and then ends the session) was not captured on their systems.

The available data included manually collected start time at Ohio State, partial start time at Dallas, and start and end indications at LC. However, even though LC had "theoretical" session boundary markers, there were no time stamps on the data for time of day, nor were there any stamps indicating the date.

For Dallas, time was recorded to the nearest minute, which is insufficiently detailed for transaction analysis because patrons enter commands more often than once per minute. At Syracuse, an attempt was made to capture session beginnings by requesting certain user input, but this was often disregarded. The use of the Ohio State data from 1981 allowed session determination by coordination with observers who recorded the beginning of sessions. Since this was the only direct observation data, it was thought to be advantageous to make additional use of it even though it was originally collected for another study. Syracuse had no user codes that were precoded. We obtained the literal use codes from the screen and had to search through the tapes until we came across the codes. System messages were also extracted in this manner.

C. METHODOLOGY

The task of analyzing the transaction tapes was broken down into the following steps:

1. Obtain the transaction tapes.
2. Check the data for validity and format; then copy them.
3. Condense the tapes by eliminating any data unnecessary to the analysis.
4. Determine sessions where possible.
5. Conduct analyses of nontransitional type (i.e., determine time and/or commands per session, etc.).
6. Build state maps using all codes.
7. Run analysis, check for problems, and condense the mapping further if required.
8. Build special maps and subsets of data.
9. Run analysis again.
10. Examine results for observations and conclusions.

The particular methodology chosen for analysis consists of first defining state codes and then developing taxonomies of these state activity codes (both user and system codes) that are mutually exclusive and exhaustive. Next, state mappings are determined that allow these state codes to be mapped into user or system states. User and system transactions are mapped into activity states. This mapping depends upon the user or system code assigned (if there is one) and also upon the context of use. The state codes represent search phases that the user is in at a given time. These were grouped according to function into one of 11 primary states, such as states representing the beginning of a session, display of bibliographic records, set creation (simple searches), help functions, combined functions, and errors that may occur.

Numerous mappings are possible for the same OPAC system. There is no one "correct" mapping; however, there may be "incorrect" mappings. For each system, a number of different mappings were made and analysis performed. At times, the output indicated the need for modifications in the selected maps. In these instances, changes were incorporated. As such, the primary mapping serves as a structure that may be applicable in varying degrees to different online catalogs.

An individual taxonomy was often further refined in a more detailed taxonomy. For example, specific display commands could be broken down into display-author, display-title, and display-subject. It is possible to break out specific error types if such error types are defined by the system under study. Similarly, certain paging commands may be associated with some specific states, but not others.

When assigning commands to specific states, the context of the search should be considered. That is, we should consider the purpose or context of its actual use. A given command may have more than one use in practice. Therefore, its intended use should be considered, in addition to the semantics of the situation, before mapping the code to a state. It may thus be possible to map a command code to more than one state, depending on the context of its use.

An example of this is shown below, using search terminology from the Library of Congress system. The third column illustrates the state mapping that would occur for an "original" map, while the right column illustrates where the code would be mapped for a "context validated" map. BGNS represents a begin session code, BRWS represents a browse code, and DISP represents a display state, while ENDS represents an end session state.

Here, the paging commands were originally mapped into a display state, since they resulted in a new screen display. But, when considering the actual context of paging command use, they were being used as a browsing command (a specific bibliographic record was not being looked for in this example). Thus, these commands were mapped to a browse state (BRWS or BRW+, which means browse with explicit paging).

SEARCH	SEARCH	STATE	
CODE		Original Map	Context-Validated Maps
BGNS	LCCC	BGNS	BGNS
BRWS	PSYCHOTHERAPY	BRWS	BRWS
ƀ	paging command	DISP	BRW+ or BRWS
ƀ	paging command	DISP	BRW+ or BRWS
ƀ	paging command	DISP	BRW+ or BRWS
ENDS		ENDS	ENDS

Once these state codes and mappings were determined, transition probability matrices and state diagrams were calculated and constructed. State transition matrices simply represent in matrix form the probability (or percentage) of going from an origin state to a destination state (e.g., the probability of starting in a display state and ending in an error state or starting in an error state and ending in a help state). The state diagrams then illustrate the frequencies (percentage) of starting in a given state and ending up in another state.

Such matrices and diagrams can be calculated for 1st, 2nd, 3rd, and higher orders, where orders are defined as the number of transitions out of the origin state. That is, a transition from display to display is 1st order, while a subject to subject to author transition is 2nd order.

D. OBSERVATIONS AND CONCLUSIONS

In attempting to draw conclusions and observations from the transaction analysis data, a number of questions were proposed and then addressed with regard to specific systems. A number of the original questions could not be answered because of the nature of the available data (e.g., time stamp not available on given systems) and data collection period. Observations and conclusions include:

The amount of time spent at the terminals was not available with any accuracy except at The Ohio State University, where it ranged from an average of approximately 4 minutes to over 9 minutes, depending on the library considered, i.e., Engineering, Undergraduate, West Campus, Education or Main Library. In all cases, the time spent at terminals was higher than the time spent at card catalogs (if available) at the same institution.

Errors frequently occurred in sequence, i.e., when an error was made, there was a high tendency to make another error (.286 at Syracuse, .598 at LC, .333 at OSU). At Dallas Public, few errors were made, a situation we attribute to the format-filled system, which allows for few syntactic errors.

A number of questions are raised by the tendency to remain in an error state, including: Would help commands be useful? Are the help commands available useful?

The data indicates that, for at least three of the systems, Library of Congress, Dallas, and The Ohio State University, we are dealing with fairly homogeneous data. We cannot say with any certainty that we have a random sample, especially at Syracuse, Dallas Public, and Library of Congress. No significant difference was found in the 0th order probabilities for The Ohio State University when using spring quarter data. Similarly, the March and June data from LC were comparable, as were the 1st and 2nd week of the Dallas Public data.

These comparisons were done on simple frequencies of command usage. It is not possible to compare state chaining patterns of higher orders with any degree of confidence. We are dealing with complex chains and behaviors that are dependent on numerous, undefined variables. Indeed, even with regression analysis techniques that are well defined, it is important to note that results are very dependent on the observed data sets. Another factor is that random samples were not obtained. The systems are not designed to provide all the required data or to select random samples for analysis. Random samples would be necessary for strict sampling validity.

For LC, Syracuse, and Ohio State, the transitions of 1st and 2nd order between identical states have high probability, usually the 1st or 2nd highest probability in the transition matrices. This would seem to imply that users have a tendency to stay or return to the same state up to a chain of three commands. In addition, a detailed analysis of 2nd orders tends to confirm that they remain in the same state throughout the chain.

Search states have a tendency to go to the same states. For example, an author search tends to proceed to another author search, a subject search tends to proceed to another subject search, etc. Users seem to remain in the same search state for a long time. As previously mentioned, this also occurs for the error state.

A look at the most frequently used search types by title, author, subject, etc., yields the following:

	OSU(%) Spring 1981	DALLAS(%) April 1982	SYRACUSE(%) April 1982
Title	30.8	16.0	20.7
Subject	24.6	60.0	26.2
Author	16.8	15.0	24.0
Author/title	10.5	3.6	3.7 (Boolean)
Other	17.3	5.4	25.4

Subject searches are highest at Dallas and Syracuse, but at Syracuse, where the subject search is the highest percentage search, more sessions begin with an author search than with a subject search (25.1% compared to 21.9%). This "enhanced" number of subject searches may be attributable to users who use the search often within a session and may not be an indication of the type of search chosen. In addition, subject searches result in zero hits more than other searches, with the exception of a DUAL search (Boolean) at Syracuse. This indicates that subject searches are not a particularly successful type of search.

E. TRANSACTION ANALYSIS GOALS

In summary, transaction log analysis is an effective tool for both formulating the questions concerning the actual user behavior at online catalogs and in answering these questions. There are a number of goals in studying this behavior, including the determination of how users perceive the system, how the system fits the users' needs, and how the system is really being used. This third goal most obviously lends itself to transaction log analysis. The users' perception of the system and how it fits their needs helps to dictate how the system is actually used.

In order to approach a transaction analysis, it is first necessary to consider the system at a "lower" zero level. Next considered are 1st order descriptions, very simple user behaviors and system responses (e.g., whether the user hits a certain key or another key, and whether the system responds with a certain message or another message). At this point, we start collapsing strings of units that are coupled with a high probability into more comprehensive behavior descriptions. For example, one could collapse all repeated search strings together, such as author-author-author or repeated error strings. Or, one could collapse into a specific search command state with hits and another search command without hits. At this stage, the analysis is rerun (i.e., recalibrated) with new coding/state mappings, and other highly coupled states are sought.

Another use of the method of transaction analysis is to test hypotheses concerning user behavior at online catalogs, e.g., to test the hypothesis that a user tends to stay within the same search--author, subject, or title--within the same session. Other hypotheses may look at error occurrences and search frequencies.

Transaction log analysis may also be utilized to conduct basic exploratory data efforts and to generate questions that we may or may not be aware of. An example of this is the finding that the highest frequency of a search type is author search, while subject search is 2nd highest (this is only an example). Examination of the data may reveal, however, that within sessions multiple use of the author search is used much more frequently. Thus, when examining the frequencies of sessions that have specific search types, subject search may be highest.

This first online public access catalog transaction log analysis study has helped to refine the goals of transaction analysis. It has also revealed some of the problems associated with conducting such a study, which should be beneficial to future efforts. It has increased our knowledge of how these systems are actually being used and provides data and results with which to formulate questions and conclusions on the operation of OPACs. As such, it offers an opportunity to build upon these efforts. Transaction log analysis offers an opportunity to expand our knowledge of the actual use of OPAC systems.

1.0 INTRODUCTION

From November 1981 to April 1983, the OCLC Office of Research conducted research into online public access catalogs (OPACs). This project was funded by the Council on Library Resources, Inc. (CLR). Like similar projects funded by CLR, this one attempts to provide new insight into the use of online catalogs by obtaining information that may serve as input for better system design of OPACs. Based on this information, OPACs could be designed to incorporate not only the desired user features, but to provide more effective searching techniques.

This first part of the study, transaction log analysis, focuses on the current utilization of online public access catalogs. The methodology was chosen to obtain machine-readable transaction logs from the online catalogs; analyze these transactions via stochastic search pattern development and mathematical model development utilizing Markov analysis; and develop transition probability matrices.

1.1 PROJECT OBJECTIVE

The objective of Volume I as stated in the proposal to the Council on Library Resources, Inc., is "to discover the extent to which current system features are used and any patterns of that use." The underlying assumptions for system design (e.g., the need for subject access or term searching, diverse display options, and help-related informational features) were compared with system usage. This comparison between system design and use is essential to improve or redesign OPAC systems and to help evaluate current system designs.

1.2 PROJECT SCOPE

Current patron use was obtained via transaction log tapes, i.e., machine-readable logs from several online public access catalog systems: Library of Congress; Syracuse University; Dallas Public Library; The Ohio State University. In addition, transactions on OCLC public terminals at 53 libraries were extracted from daily transaction logs, giving frequencies of search types. These include 5 public libraries, 5 federal or state libraries, and 43 academic libraries (including several medical, health, and law libraries). These systems represent academic institutions, a national library, and a public library, providing us with a diversity of operating OPACs. At the macroscopic level, statistical measures have established relative use of system and user commands; at the microscopic level, sequences of actions at terminals have established individual user patterns.

2.0 DATA COLLECTION

Current patron use patterns were obtained via transaction tapes from the online public access catalogs for the following time periods:

Library of Congress - March to June 1982
Syracuse University - April 12 to April 28, 1982
Dallas Public Library - August 9 to August 21, 1982
The Ohio State University - January to December 1981, March to May 1982, and November 1982
OCLC Public Terminals - May 1982 to October 1982

The institutions operating the systems under study selected the data collection date ranges, which varied from two weeks to one year. The amount of data collected was sizable, both in absolute terms and in effort required for reduction and analysis. The analysis of Syracuse University and Dallas Public Library involved the entire data set. It also included an examination of the representativeness of subsets of data. For the Library of Congress and The Ohio State University, specific subsets were analyzed as explained in detail later in this section.

Each OPAC system is different; consequently, the resulting transaction log tapes varied both in content and format. These differences, which are more than minor, represent a major effort in data reduction and analysis effort. It was necessary to develop separate algorithms and corresponding computer programs to accomplish the task of data reduction, formatting, and preparation for transaction analysis.

Specific data elements collected included:

1. Terminal identification
2. System used (if more than one available)
3. File searched
4. User command used
5. Search key used
6. Beginning time of session
7. Ending time of session
8. System response
9. Time of system response
10. Search by text
11. Date

Not all of these data elements were available for each OPAC because the systems were not designed for research. Both the cost in time and the cost of hardware or software development for the purpose of recording data needs to be taken into consideration by institutions when developing data recording capabilities.

Table 1 summarizes the empirical transaction data log elements for the Library of Congress, The Ohio State University, Syracuse University, and the Dallas Public Library.

Table 1. Summary of Empirical Data Log Elements

LIBRARY OF CONGRESS	OHIO STATE UNIVERSITY	SYRACUSE UNIVERSITY	DALLAS PUBLIC
Collected 1. Terminal ID 2. System (SCORPIO or MUMS) 3. File searched 4. Command 5. Search key 6. Begins 7. Ends	Collected 1. Terminal ID 2. Start time (enter command) 3. Time of system response 4. Date 5. System response (return code) 6. Search key text	Collected 1. Terminal ID 2. Date 3. Actual command 4. System Response	Collected 1. Terminal ID and type 2. User commands and system responses (partially coded) 3. Date 4. Time (in minutes) 5. "Partial" begin session
________	Observed 1. Patron start time 2. Sex of patron 3. Patron or staff 4. Help received from staff	________	________
Not Collected 1. Time 2. (Date) 3. System response 4. # of hits 5. Patron information	Not Collected 1. Number of matches retrieved by search 2. User ID 3. Session end (patron end time)	Not Collected 1. Begin Session 2. End Session 3. User ID	Not Collected 1. User ID 2. Some user inputs? 3. All "begin" session 4. Session end
Calculated 1. State frequency distribution 2. State chaining 3. Chain length and patron frequency distributions 4. Transition probability matrices 5. Theoretical frequencies 6. Sessions 7. Session length in number of commands per session	Calculated 1. Elapsed time between transactions "gap time" or "command time" 2. Session length i.e., number of commands per session 3. Partial "session" length in elapsed time 4. "Sessions" 5. Relative frequency of return codes 6. Card vs. LCS use 7. Percent of patrons obtaining help 8. # errors/# commands for ith command in aggregate session 9. Return code by command 10. State frequency distributions 11. Chain length and pattern frequency distribution	Calculated 1. State frequency distribution 2. State chaining 3. Chain length and pattern frequency distributions 4. Transition probability matrices 5. Theoretical frequencies 6. Time/session 7. Commands/session 8. Number of hits/search type 9. Error types (frequencies) 10. Sessions	Calculated 1. Frequency of search commands in a session 2. Sessions 3. Number of terminals used each day 4. Transition frequencies 5. Chain length & pattern 6. Theoretical frequencies 7. Session length (time)

2.1 LIBRARY OF CONGRESS

The Library of Congress system is actually two systems--SCORPIO (Subject Content Oriented Retriever for Processing Information Online) and MUMS (Multiple Use MARC System). SCORPIO is used by both the public and the staff, while MUMS is basically staff-oriented.

Data was obtained for public users of SCORPIO (i.e., a "public" terminal) during the months of March 1982 and June 1982, for library staff users during April 1982, and for both public and staff users during May 1982. Characteristics of the tape data include user input command, no system response code, beginning-of-session marker, end-of-session marker (if the user chose to use such), and terminal identification.

Unavailable data elements included time stamps, date stamps, system responses, number of hits, and patron identification.

Only the data from SCORPIO and public users was utilized for this project; the entire two months of data was evaluated. This study was interested in the public users so the analysis of MUMS would have been essentially another system. In addition, MUMS data had a number of problems with the actual tapes.

2.2 SYRACUSE UNIVERSITY

The transaction log data received was collected during two weeks in April 1982. Specific data collected for each transaction terminal includes identification (public use terminals only), date, user command, and system response. Items not available included both beginning- and end-of-session markers and user identification. Although an attempt was made to obtain the beginning of sessions, the users had a tendency to ignore it. It would appear that asking patrons to cooperate in this respect is not useful.

2.3 DALLAS PUBLIC LIBRARY

The data collected includes both public and staff terminals from August 9 to August 21, 1982. No user identification, session beginnings, or session ends were available; some user inputs were also unavailable. Terminal identifier and type, user commands, partially coded system responses, dates, and times (to the nearest minute) were available.

2.4 THE OHIO STATE UNIVERSITY

The most extensive transaction log data was obtained from The Ohio State University, since we had access to transaction tapes for the whole of 1981. This data was collected for a previous study for the National Science Foundation on terminal requirements for online catalogs. We decided to take advantage of this and to examine any pattern differences during the academic year, including the summer quarter. The 1982 data included a change in some command structure and was not utilized.

Data elements available from Ohio State included terminal identifier, starting time (for entering a command), user command, time of system response (return code), and search text. User identification and session end were not collected. In addition to the data from the transaction tapes, data for certain time periods (when observers were stationed at the terminals of five different campus libraries) was available. This data included patron start times, patron sex (which helped data collectors determine new users as they arrived), patron or staff, and help received from staff.

Table 1 summarizes the empirical data elements for all four OPAC systems. It lists the data elements collected, not collected (not available), externally observed (Ohio State), and calculated for each system (explained in section 3, Methodology).

2.5 OCLC PUBLIC TERMINALS

In mid-year 1982, approximately 2,200 libraries had OCLC terminals, enabling them to use one or more of the four major subsystems of the OCLC Online System (Cataloging, Interlibrary Loan, Serials Control, and Acquisitions). In addition, more than 700 other libraries used dial-access terminals to access OCLC through TYMNET or direct dial. Some libraries share terminals with one or more other libraries, some are customers of processing centers, and some OCLC member libraries have several terminals. In all, over 5,000 OCLC terminals were linked to the OCLC Online System in 1982. Of the more than 3,300 OCLC member libraries, nearly 60% were academic libraries.

Although most OCLC terminals are used primarily to perform the technical or "housekeeping" functions supported by one or more of the four subsystems, a few libraries have chosen to place OCLC terminals in public areas for use by patrons and public service staff. This option provides staff and patrons online access to bibliographic and location information for the more than 9 million records in the Online Union Catalog. Patrons may retrieve bibliographic records from the Online Union Catalog by typing simple search commands at an OCLC terminal keyboard. Additional commands permit the user to identify through location symbols the libraries that have used a record for cataloging an item in their collection. Each bibliographic record in the OCLC database averages 13 location symbols.

While OCLC public terminals provide online search access for patrons and public service staff, many of these terminals are multipurpose in use. A terminal may be logged on to the Interlibrary Loan or Serials Control Subsystems, for example, and used by staff to perform the functions of those subsystems. These activities may be separately scheduled, or performed alternately with patron bibliographic searching. Searching of the bibliographic database is possible in any of the four subsystems.

In September 1982, 76 OCLC terminals were used as public terminals at 53 different libraries (as defined by unique OCLC location symbols). By institution type, these libraries included 5 public libraries, 5 federal or state libraries, and 43 academic libraries (including several medical, health, and law libraries).

3.0. METHODOLOGY

3.1 PROBLEM STRUCTURE

The general methodology was structured as follows:

1. Obtain the transaction tapes (machine-readable logs).
2. Reduce these via reformatting to more usable form.
3. Develop a taxonomy of state activity codes that are mutually exclusive and exhaustive.
4. Develop system state mappings, including system codes and search commands.
5. Break down the logs into user sessions, if possible (the parsing of the logs and assignment of state activity codes).
6. Subsequent analysis of the data via development of transition probability matrices (TPMs), pattern diagrams, Markov chain analysis, and probability distribution models.

Figure 1 illustrates the steps performed in the transaction analysis methodology.

These steps were performed for each OPAC system studied. Since each was different, its corresponding transaction log was different--both in content and format. Consequently, completely separate algorithms and corresponding computer programming were required to accomplish steps 2 through 5 in the analysis methodology. In fact, the breaking down and reformatting of transaction logs, parsing into user sessions, and development of state codes are a major part of the analysis, considering man-hours of work and computer analysis requirements. At this stage in the development of transaction analysis, it does not seem likely that the situation will change quickly.

Future data collection efforts could be facilitated by clearly defining the production requirements of data log tapes before the system is put into operation. This was impossible to do with the data we acquired because the systems were not designed for research or to provide this level of information.

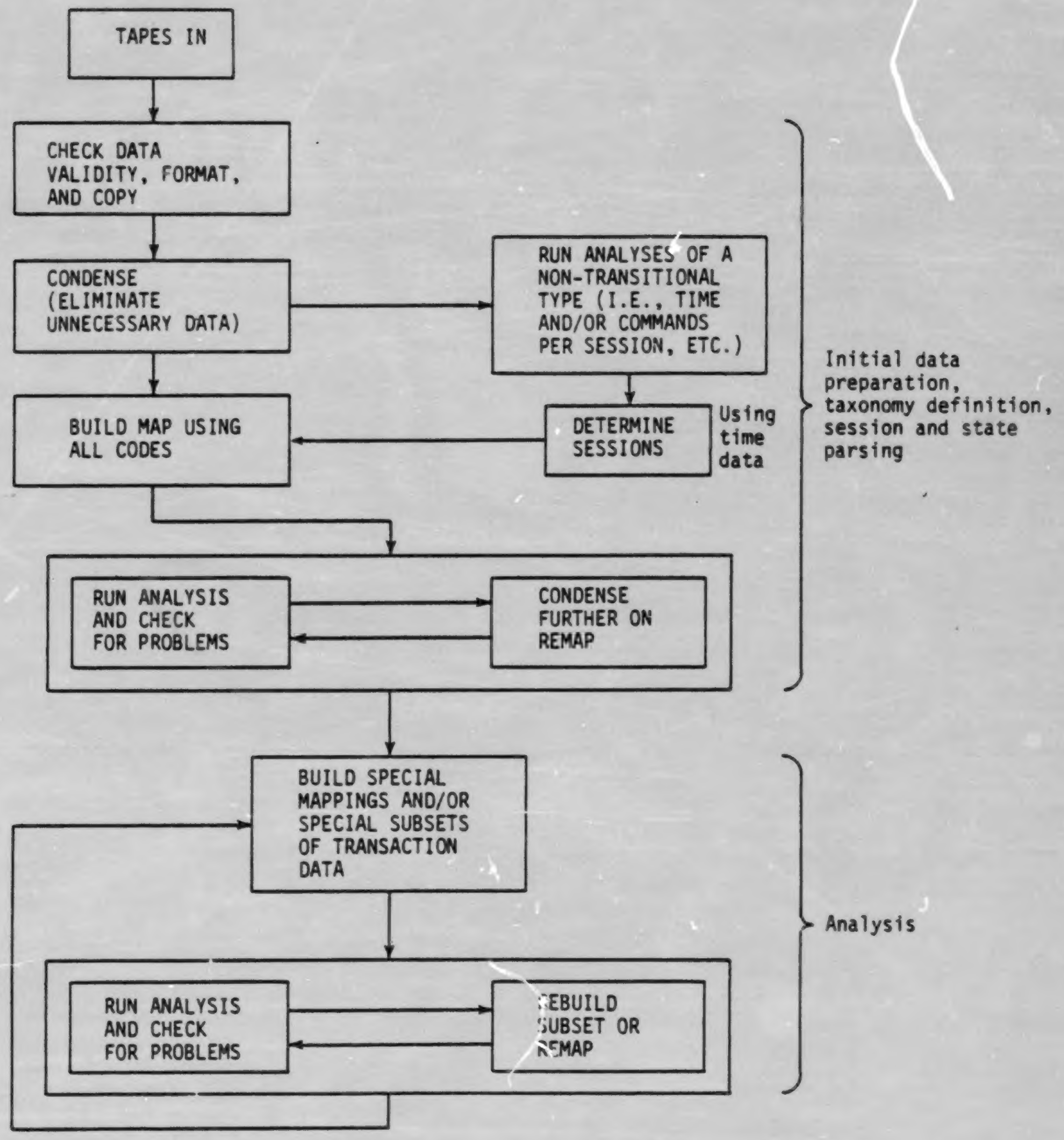
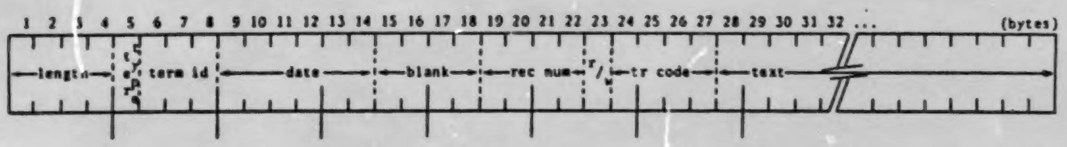

TAPES IN
CHECK DATA VALIDITY, FORMAT, AND COPY
CONDENSE (ELIMINATE UNNECESSARY DATA)
RUN ANALYSES OF A NON-TRANSITIONAL TYPE (I.E., TIME AND/OR COMMANDS PER SESSION, ETC.)
BUILD MAP USING ALL CODES
DETERMINE SESSIONS
Using time data
RUN ANALYSIS AND CHECK FOR PROBLEMS
CONDENSE FURTHER ON REMAP
Initial data preparation, taxonomy definition, session and state parsing
BUILD SPECIAL MAPPINGS AND/OR SPECIAL SUBSETS OF TRANSACTION DATA
RUN ANALYSIS AND CHECK FOR PROBLEMS
REBUILD SUBSET OR REMAP
Analysis

Figure 1. Flow Diagram for Transaction Analysis Methodology

3.2 TRANSACTION TAPE REDUCTION AND PARSING: SESSION DETERMINATION

The transaction log tapes varied greatly in both content and format and thus required both a reduction in information and a normalization (where possible) of the information to allow analysis. Figure 2 illustrates the transaction log record format chosen to arrange the data. As such, it represents step 2 in the general methodology. In some systems, a "begin session" command was available (e.g., Library of Congress, The Ohio State University), while others had none. Similarly, an "end session" command was not available in any system except the Library of Congress, but users were not required to use it. Clearly, such limited session boundaries posed extreme problems in determining session length and the subsequent parsing of data into user sessions.



length - contains the integer number of bytes in that log record.

ter type - a single digit integer representing the type of terminal.
(1 = public, 2 = staff, 3 = both public and staff.)

term id - three digit integer field representing the terminal identification number.

date - the date for this transaction in YYMMDD format (if known).

blank - four bytes of blank filler (for compatability with other systems).

rec num - sequential record number for terminal for that day.

r/w - integer indicating whether record was read from the terminal or written to the terminal. Library of Congress data includes only user input information, thus all records will be read from the terminal (value = 1).

tr code - corresponds to the four character transaction code associated with this particular screen as generated by Library of Congress subroutines.

text - the byte for byte representation of what was read from the screen.

Figure 2. Transaction Log Record Format

3.3 TAXONOMY OF STATE SELECTION, SYSTEM SEARCH COMMANDS, AND STATE MAPPINGS

Before proceeding with the analysis, we had to develop a definition of the user and system commands and to classify these commands into state codes. The state codes represent search phases that the user is in at a given time. Search commands were grouped according to function into one of 11 primary states as defined in Figure 3.

In defining a state command mapping, there is no one "correct" map. State definitions are open for discussion, especially when applied to a specific system. The functional classification of states shown in Figure 3 was thoroughly discussed among the researchers at OCLC. The classification was also submitted for comments to individuals at the institutions supplying transaction tapes. The primary mapping provides an overall structure which may be applicable to different online catalogs in varying degrees.

The process function used is listed along with a state map code consisting of four letters. Such processes are not necessarily followed in order, nor are all the states available to each system (e.g., a help function). Explanations of specific functions for individual OPAC systems are contained in section 4.0.

Two notes must be made at this point. First, a primary mapping, such as that shown in Figure 3, may be too general. In such a case, a nested taxonomy of states is appropriate (if possible). Such a taxonomy consists of the primary mapping with specific breakouts of state codes in greater detail (e.g., with display commands or errors broken down into specific types).

The second note concerns "context validation," a term coined by Charles Hildreth, OCLC Research Scientist. Context validation is concerned with assigning specific commands to state codes in the context of a search. That is, considering the purpose or context of its actual use, a given command may have more than one role, so that we must know the purpose of its use before assigning a state mapping to the command code. Thus, it may be possible to have a "one to many" mapping of a command code to different states, depending on its context. In most cases, this context may be determined by checking the preceding different command before state assignment.

PROCESS	STATE MAP
1. File Selection	BGNS
2. Term selection/review No sets created, this is an index review No display of records operates only within an index or thesarus (by definition)	BRWS
3. Create sets (simple search)	SRCH
4. Reformulate, negotiate, refine query (combine, limit, etc.) where you have actually done a search of the data base	MODS
5. Display records/citations	DISP
6. Review search history/search status obtaining information about the history and/or status of the session we are in View system/session information cost show settings etc.	STAT
7. A help function with its own command	HELP
8. End session cost might also be here	ENDS
9. COMBINED FUNCTIONS: FIND = BRWS/SLCT/(COMB)/DSPL SCAN = SLCT/DSPL	FIND
10. COMBINED FUNCTIONS: RETR = SLCT/COMB	RETR
11. ERROR	ERR

Figure 3. Primary Mapping for State Selection for General Use

3.4 ANALYSIS METHODOLOGIES

Once the data has been broken into state codes, the development of state occurrence frequency distributions can proceed along with the development of state transition chains or strings. Sequences of states, i.e., state strings, are created providing both individual frequencies on state occurrences and frequencies on higher order state and state string occurrences. These represent observed data distributions. They may be used for comparison with the theoretical state string predictions via Markov analysis.

3.4.1 Markov Analysis of State Chains

In order to understand the higher order analysis of state occurrences and state-string occurrences, it is necessary to consider briefly the application of conditional probability and Markov chain analysis.

Suppose we let S_i represent the state i in our classification of possible states into which online search codes may be mapped. Allow the outcome of any transaction, i.e., being in state i, to depend on the outcome of the directly preceding state. Suppose further that P_i represents the probability of the ith state occurring. This P_i then may be associated with the observed frequencies of occurrence of the state i and is considered to be the 0th order transition. It may be obtained simply by letting $P_i = O_i/n$, where O_i is the observed frequency of ocurrence of the ith state, and n is the total number of observed transaction occurrences (of all possible states).

Then S_i may be associated with P_i for the 0th order frequencies by the relation:

(Eq. 1.) $p_i = P(S_i)$ $\left[\text{0th order}\right]$; p_i = probability of ith state occurrence

Now, if the outcome of any transaction, i.e., being in state i, depends on the outcome of the directly preceding state, then S_i is no longer associated with a fixed probability P_i; but to every pair (S_j, S_k) there corresponds a conditional probability, P_{jk} such that, given that S_j occurred at some step (transaction), the probability of S_k occurring at the next step (transaction) is P_{jk}.

(Eq. 2.) $p_{jk} = P(S_k|S_j)$ $\left[\text{1st order}\right]$

In addition to the P_{jk}, we must be given the probability P_k of the outcome state S_k at the initial step, i.e., P_k = 0th order probability of S_k. Then the probabilities of sequences or "chains" occurring following the initial state corresponding to 2, 3, and 4 states (2nd, 3rd, and 4th order) must be defined as:

(Eq. 3.) $P\{(S_j,S_k)\} = p_j p_{jk}$

(Eq. 4.) $P\{(S_j,S_k,S_r)\} = p_j p_{jk} p_{kr}$

(Eq. 5.) $P\{(S_j,S_k,S_r,S_t)\} = p_j p_{jk} p_{kr} p_{rt}$

and in general:

(Eq. 6.) $P\{(S_{j_0},S_{j_1},S_{j_2},\ldots,S_{j_n})\} = p_{j_0}\, p_{j_0 j_1}\, p_{j_1 j_2} \cdots p_{j_{n-1} j_n}$

The initial state being number zero, the 2nd state number 1, etc., where:

(Eq. 7.) p_{jk} = probability of transition from state S_j to state S_k

Such outcomes define a Markov chain of transition from one state to another state.

3.4.2 Transition Probability Matrices (TPMs)

We may arrange the probabilities P_{jk} defined in Equation 7 in a transition probability matrix (TPM), P, as shown below:

(Eq. 8.) $$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & & \ddots & \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

Here P is a square stochastic matrix; the P_{ik} 's are non-negative probabilities; and $\sum_{k=1}^{n} P_{ik}=1$ for all i. P, together with the initial distribution $[P_k]$, completely defines a Markov chain with states S_1, S_2, ---, S_n.

A transition from S_j to S_k may occur in n steps by different paths, such as S_j --- S_{j1} --- --- --- S_{jn} --- S_k. Then the conditional probability that the system takes this particular path, starting at S_j, is given by:

(Eq. 9.) $P(S_j \to S_{j_1} \to \ldots \to S_{j_{n-1}} \to S_k | S_j) = p_{j j_1}\, p_{j_1 j_2} \cdots p_{j_{n-1}}\, p_k$

The sum of all possible paths from state j to state k may be denoted by the probability P^n_{jk} , where the superscript n refers to n stops in traversing the path. We then have:

(Eq. 10.) $p^n_{jk} = \sum_{\text{all } i} p_{ji} p^{n-1}_{ik} \quad ; \quad n \geq 1$

We may also create a stochastic transition probability matrix of the P^n_{jk} and call it P^n, as shown in Equation 4.11. For n = 1, $P^n = P$, as shown in Equation 8.

(Eq. 11.) $$P^n = \begin{bmatrix} p^n_{11} & p^n_{12} & \cdots & p^n_{1n} \\ p^n_{21} & p^n_{22} & \cdots & p^n_{2n} \\ \vdots & & \ddots & \\ p^n_{n1} & p^n_{n2} & \cdots & p^n_{nn} \end{bmatrix}$$

Then, using Equation 10. to find P^n_{jk} , we multiply the elements of the jth row of P by the elements of the kth column of P^{n-1} and add all the products, i.e., $P^n = P^n P^{n-1}$.

In this manner, we may create our higher-order transition probability matrices.

3.4.3 Kolmogorov-Smirnov Test

Once these theoretical, higher-order transition probabilities and matrices have been developed, we may also calculate, dependent upon the 0th order frequencies, the observed probability transitions from one state to another. The limitations on any calculation, observed or theoretical, are the number of states considered and the order desired, since possible combinations increase geometrically with states and order. The Kolmogorov-Smirnov test is used to compare these distributions for correctness of fit, which simply considers the maximum deviation of the sample cumulative distribution function F_n from the theoretical cumulative distribution function F, for a sample size n. That is, let:

(Eq. 12.) $D_n = \max | F_n - F |$

D_n has a distribution that may be found in tables such that we may determine:

(Eq. 13.) $P(D_n > t_\alpha | F \text{ is the true distribution}) = \alpha$

Then if $D_n > t$ we reject F at significance level α

This essentially puts a boundary around the deviation of the cumulative differences of the functions. An important point should be made here: the model used may be rejected by certain individual differences but still be quite useful, especially as a simulation model. The data must be examined and not rejected simply by the nonfit at a given significance level.

4.0 RESULTS

The methodology defined in the previous section was applied to the selected OPAC systems. Transitions from one state to another were determined after the reduction and parsing of the "raw" data transaction tapes. This required the choice of state taxonomy and/or development of appropriate state mappings to fit into the primary and nested-primary mappings with context validation. Once these mappings were determined, the determination of 0th-order-observed frequencies of the state occurrences was calculated via appropriate computer programs.

The computer programming effort to accomplish the data reduction, parsing, and analysis was and remains significant, both because of the complexity of the tasks required to develop logic and specific algorithms and because of the enormous amounts of data.

The generation of higher-order-state-transition frequencies was also achieved via computer algorithms. At this stage, theoretical state transition probabilities were calculated by means of Markov chaining analysis programs for numerous state mappings, both in combinations and for the different systems. In fact, it was not until complete mappings were made for each system and transition probabilities determined, examined, and thoroughly discussed, that the concept of "context validation" occurred. It was at this time that it became evident that it was necessary to look at more than the semantic meaning of the user command in question, since the same command may have more than one purpose in actual use (i.e., one to many mappings). In this sense, the "actual" use of the command must be considered when mapping a particular command (code) to a state. The difference was determined between functional and semantic command usage in this manner.

The statistical testing of the observed and theoretical state chains was then done and transition probability matrices (TPMs) developed, along with state transition diagrams; a series of hypotheses/questions was proposed and examined to determine if each could be addressed satisfactorily.

4.1 PROPOSED HYPOTHESES/QUESTIONS

The following is a list of questions that may be considered. These are not intended to be conclusive, but they serve as typical questions that may be considered. Each of these was considered for the various systems.

1. How much time did patrons spend at the terminal?

2. What commands did they use?

3. How did time per transaction vary by command?

4. What errors did people make? Where in the session did they make them?
5. How does behavior vary according to session length?
6. How does use vary among libraries (within the same institution)?
7. How does use vary over time (such as academic quarters)? Within libraries (in the same system)?
8. What is the frequency of "hits"/no "hits"?
9. Given that the searcher is in a search type, such as author search, what is the likelihood of changing search type?
10. Are there sophisticated searching features available that are used infrequently? What correlation, if any, does type of library have with needed prompting devices or search sophistication?
11. What is the relative frequency of available fields searched?
12. What is the relationship between rigid command syntax, command usage, and error message frequency?
13. What, if any, patterns of feature use exist?
14. How does user time at the terminal compare with each type of search (e.g., author vs. subject)?
15. If an error is made, what is the likelihood of continuing to make an error?
16. Is there any relationship between errors and length of search?
17. What are the frequencies of single-type search sessions and multiple-type search sessions?

4.2 LIBRARY OF CONGRESS

Table 2 illustrates the Library of Congress SCORPIO commands and primary state mapping with context validation. In addition to this primary mapping, another mapping was made for the Library of Congress that broke down the mapping to explicit paging commands in addition to the primary mapping.

Table 2. Primary Mapping SCORPIO - LC States/Commands

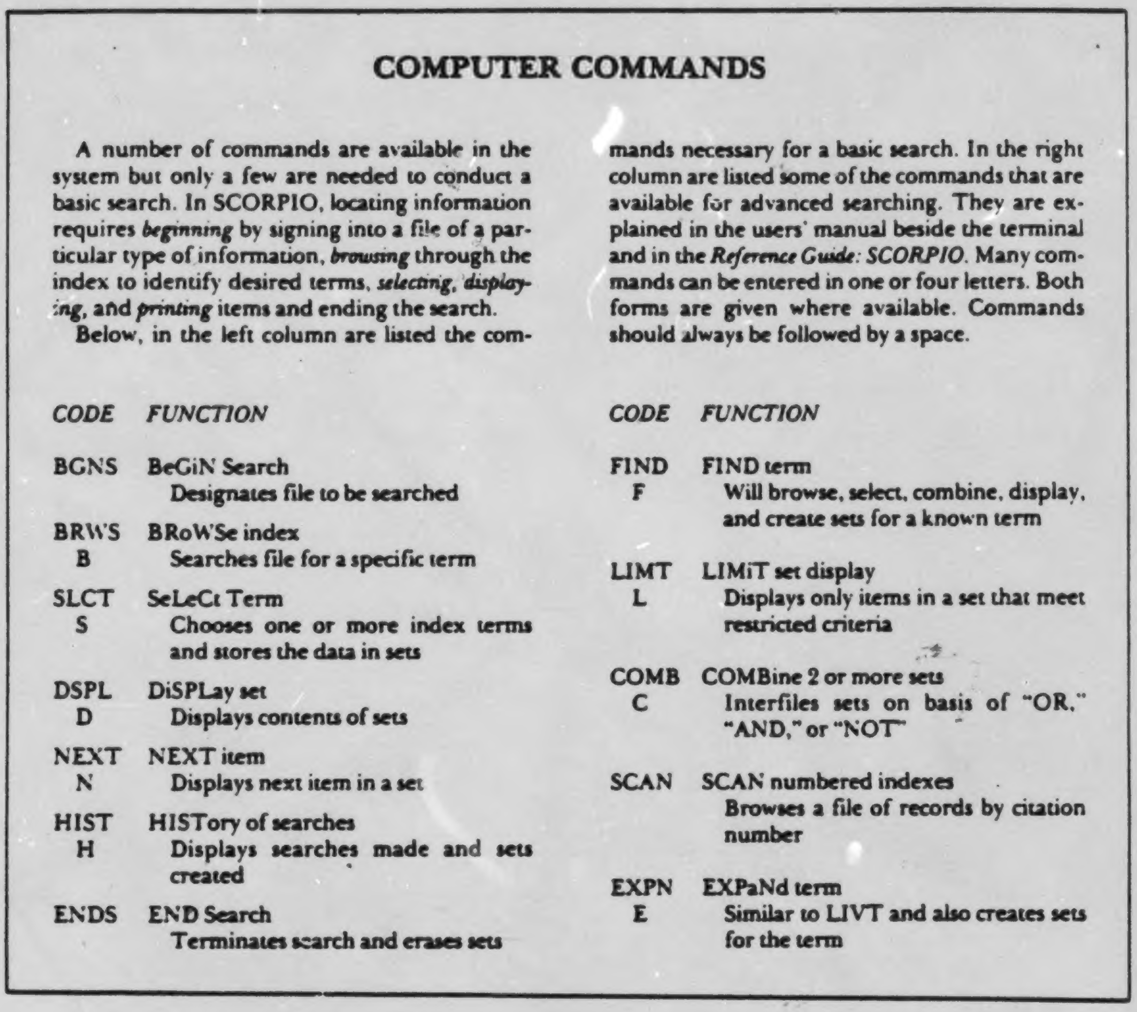
PROCESS	STATE MAP	LC COMMAND
1. File Selection	BGNS	BGNS
2. Term selection/review	BRWS	BRWS
No sets created, this is an index review		LIVT
No display of records operates only within an index or thesarus (by definition)		
3. Create sets (simple search)	SRCH	SLCT
		EXPN
4. Reformulate, negotiate, refine query	MODS	COMB
(combine, limit, etc.) where you have		LIMT
actually done a search of the data base		
5. Display records/citations	DISP	DSPL
		NEXT
		"XMIT"
6. Review search history/search status	STAT	HIST
obtaining information about the history		RLSE
and/or status of the session we are in		SHOW
View system/session information		NEWS
cost		SET
show settings		
etc.		
7. A help function with its own command	HELP	No Command
8. End session	ENDS	ENDS
cost might also be here		
9. COMBINED FUNCTIONS:		
FIND = BRWS/SLCT/(COMB)/DSPL	FIND	FIND
SCAN = SLCT/DSPL		SCAN
10. COMBINED FUNCTIONS:		
RETR = SLCT/COMB	RETR	RETR
11. ERROR	ERR	Not a valid command

Table 3 shows some of the user commands available in the SCORPIO system along with the formal definition of the code purpose. Table 4 shows a sample search performed on the LC catalog demonstrating a search to identify books about solar houses. Line 1 indicates that the search is to begin with the command BGNS, followed by LCCC (Library Computer Catalog Center file). Line 2 shows that the patron typed the command "BRWS" followed by SOLAR HOUSES. The command "BRWS" browses and searches for a specific term. The result is a listing of 10 possible entries, labeled B01 to B10. Line 3 then indicates that the command "SLCT" selects the term B6 and stores the data in sets. Line 4 shows the display command "DSPL" followed by set #2, which results in a display of the contents of the set. Next, line 5 is the end of the search command "ENDS".

The data collected from the Library of Congress had several limitations. The documentation of the tape contents and system was not as complete as we may have wished, partly because so many options were available (SCORPIO and MUMS). In addition, there were numerous overwrites and recurring output loops within the MUMS data. The contents of two patron-only tapes were too unreliable for use; one tape consisted almost entirely of BGNS commands, and the other had instances scattered throughout of unreasonable repetitions of single commands. The remaining patron data had to be hand checked, and the obviously unreliable data taken out before further analysis could begin. Examples of rejected data included partial sessions at the end of a day or before a system crash, sessions with unreasonable repetitions of single commands, sessions of unreasonable length and contents, etc. In all cases, entire sessions were deleted, rather than individual commands.

SCORPIO data did not have either time stamps on the commands or indication of system response. Without time stamps, there is very little that can be done to note session boundaries. Judging from the contents of some sessions (defined as all searches between consecutive ENDS commands), it seems some patrons may have signed off and signed back on after each search. (Two session boundaries are BGNS and ENDS. One can either use BGNS to begin a session or ENDS to end one. There is sometimes more than one BGNS associated with a particular ENDS. Either choice introduces its own problems. We chose ENDS to avoid breaking apart individual sessions.) Conversely, because of radically different patterns and subject material searched during a "session," it appears others may have left their terminals without signing off, making it possible for different patrons to continue these "sessions." Therefore, no time usage analysis can be done. Having no indication of system response limits the type of analysis that can be done on the data. Furthermore, we could not determine if an indication had been received signaling whether or not a given command was recognizable. Such an indication would be useful.

Table 3. User Commands Available in SCORPIO - LC
(Source: Library of Congress)



COMPUTER COMMANDS

A number of commands are available in the system but only a few are needed to conduct a basic search. In SCORPIO, locating information requires *beginning* by signing into a file of a particular type of information, *browsing* through the index to identify desired terms, *selecting*, *displaying*, and *printing* items and ending the search.

Below, in the left column are listed the commands necessary for a basic search. In the right column are listed some of the commands that are available for advanced searching. They are explained in the users' manual beside the terminal and in the *Reference Guide: SCORPIO*. Many commands can be entered in one or four letters. Both forms are given where available. Commands should always be followed by a space.

CODE	*FUNCTION*
BGNS	BeGiN Search Designates file to be searched
BRWS B	BRoWSe index Searches file for a specific term
SLCT S	SeLeCt Term Chooses one or more index terms and stores the data in sets
DSPL D	DiSPLay set Displays contents of sets
NEXT N	NEXT item Displays next item in a set
HIST H	HISTory of searches Displays searches made and sets created
ENDS	END Search Terminates search and erases sets

CODE	*FUNCTION*
FIND F	FIND term Will browse, select, combine, display, and create sets for a known term
LIMT L	LIMiT set display Displays only items in a set that meet restricted criteria
COMB C	COMBine 2 or more sets Interfiles sets on basis of "OR," "AND," or "NOT"
SCAN	SCAN numbered indexes Browses a file of records by citation number
EXPN E	EXPaNd term Similar to LIVT and also creates sets for the term

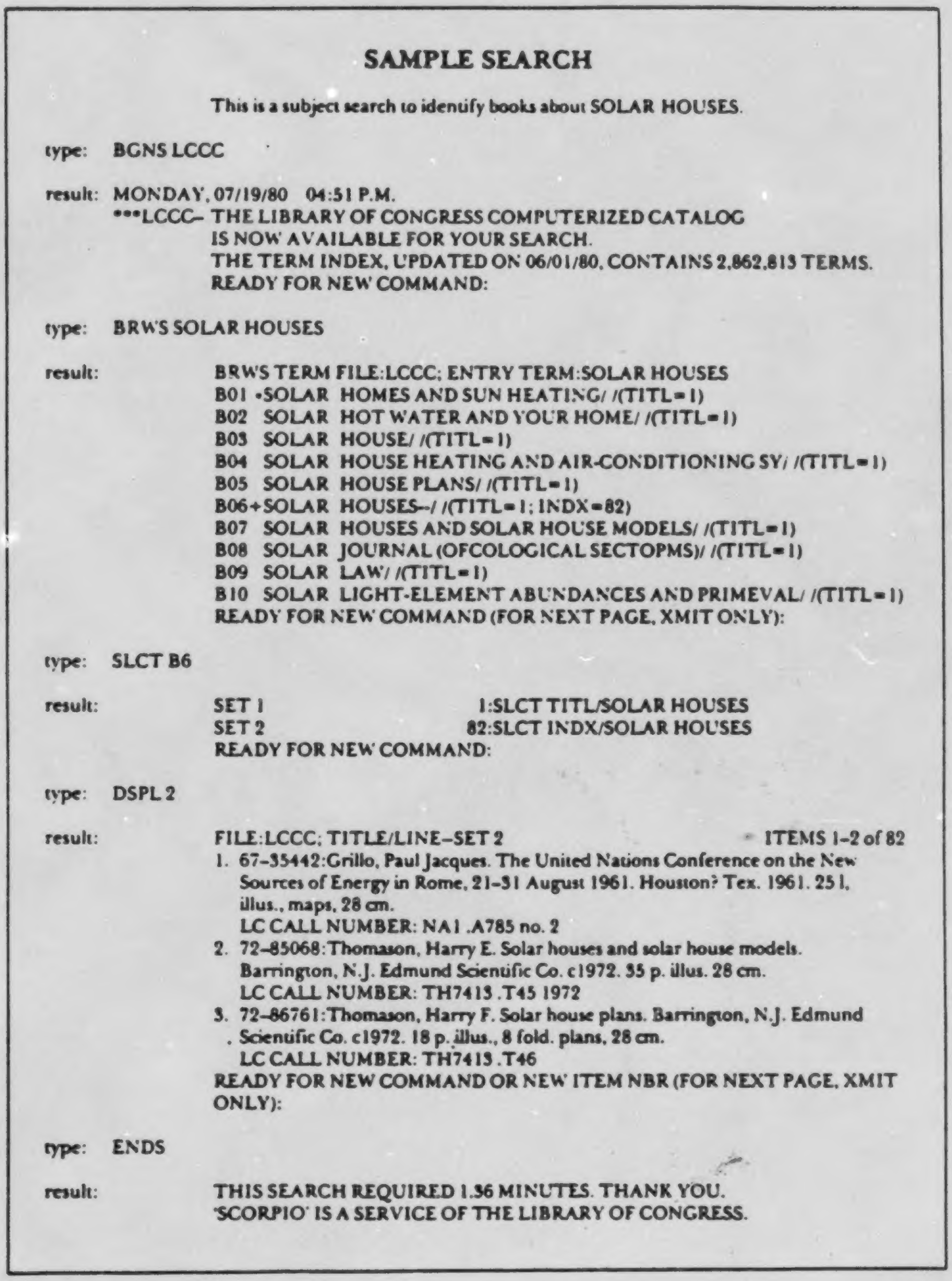


Table 4. Sample Search in SCORPIO - LC

SAMPLE SEARCH

This is a subject search to identify books about SOLAR HOUSES.

```
type:   BGNS LCCC

result: MONDAY, 07/19/80   04:51 P.M.
        ***LCCC– THE LIBRARY OF CONGRESS COMPUTERIZED CATALOG
                 IS NOW AVAILABLE FOR YOUR SEARCH.
                 THE TERM INDEX, UPDATED ON 06/01/80, CONTAINS 2,862,813 TERMS.
                 READY FOR NEW COMMAND:

type:   BRWS SOLAR HOUSES

result:          BRWS TERM FILE:LCCC; ENTRY TERM:SOLAR HOUSES
                 B01 •SOLAR  HOMES AND SUN HEATING/ /(TITL=1)
                 B02  SOLAR  HOT WATER AND YOUR HOME/ /(TITL=1)
                 B03  SOLAR  HOUSE/ /(TITL=1)
                 B04  SOLAR  HOUSE HEATING AND AIR-CONDITIONING SY/ /(TITL=1)
                 B05  SOLAR  HOUSE PLANS/ /(TITL=1)
                 B06+SOLAR  HOUSES--/ /(TITL=1; INDX=82)
                 B07  SOLAR  HOUSES AND SOLAR HOUSE MODELS/ /(TITL=1)
                 B08  SOLAR  JOURNAL (OFCOLOGICAL SECTOPMS)/ /(TITL=1)
                 B09  SOLAR  LAW/ /(TITL=1)
                 B10  SOLAR  LIGHT-ELEMENT ABUNDANCES AND PRIMEVAL/ /(TITL=1)
                 READY FOR NEW COMMAND (FOR NEXT PAGE, XMIT ONLY):

type:   SLCT B6

result:          SET 1                          1:SLCT TITL/SOLAR HOUSES
                 SET 2                         82:SLCT INDX/SOLAR HOUSES
                 READY FOR NEW COMMAND:

type:   DSPL 2

result:          FILE:LCCC; TITLE/LINE–SET 2                     ITEMS 1–2 of 82
                 1.  67–35442:Grillo, Paul Jacques. The United Nations Conference on the New
                     Sources of Energy in Rome, 21–31 August 1961. Houston? Tex. 1961. 25 l,
                     illus., maps, 28 cm.
                     LC CALL NUMBER: NA1 .A785 no. 2
                 2.  72–85068:Thomason, Harry E. Solar houses and solar house models.
                     Barrington, N.J. Edmund Scientific Co. c1972. 35 p. illus. 28 cm.
                     LC CALL NUMBER: TH7413 .T45 1972
                 3.  72–86761:Thomason, Harry F. Solar house plans. Barrington, N.J. Edmund
                   . Scientific Co. c1972. 18 p. illus., 8 fold. plans, 28 cm.
                     LC CALL NUMBER: TH7413 .T46
                 READY FOR NEW COMMAND OR NEW ITEM NBR (FOR NEXT PAGE, XMIT
                 ONLY):

type:   ENDS

result:          THIS SEARCH REQUIRED 1.36 MINUTES. THANK YOU.
                 'SCORPIO' IS A SERVICE OF THE LIBRARY OF CONGRESS.
```

4.2.1 Transition Probability Matrices, State Diagrams, and Transaction Statistics

The state transition probabilities were calculated for 1st, 2nd, 3rd, and 4th order transitions along with the 0th order state frequencies. The zero-order analysis consisted of only the frequency of occurrences of individual states that are considered independent and have no dependence on previous state occurrences. First-order analysis consisted of two-state strings and considered a conditional effect of occurrences, dependent upon prior state occurrences. Transition probability matrices were then calculated and are shown in Tables 5-12. State transition diagrams are shown in Figures 4-13. Tables 8 and 9 reflect individual March and June 1982 first order transitions. All the other tables are based on the entire set of data from both March and June 1982.

Table 5. State Frequency Distribution - Primary Mapping 0th Order - Context Validated - LC

	TOTAL NUMBER	PERCENT	JUNE NUMBER	PERCENT	MARCH NUMBER	PERCENT
BGNS	789	6	268	6	521	6
BRWS	2652	21	856	20	1796	21
SRCH	1151	10	505	12	646	8
DISP	5361	42	1940	45	3421	41
FIND	1014	8	271	6	743	9
RETR	18	1	15	1	3	3
MODS	123	1	18	1	105	3
STAT	149	1	25	1	124	3
ENDS	591	5	219	5	371	4
ERRS	913	7	216	5	697	8
TOTAL	12761	100	4333	100	8428	100

Table 6. Transition Probability Matrix Primary Mapping - 1st Order - Context Validated - Successor - LC March-June 1982 (Expressed in Percent)

	BGNS	BRWS	SRCH	DISP	FIND	RETR	MODS	STAT	ENDS	ERRS
BGNS	15.5	64.3	0.0	1.1	1.9	0.1	0.1	0.3	9.4	6.7
BRWS	1.0	48.3	28.3	4.4	8.9	0.3	0.1	0.6	4.8	3.4
SRCH	0.4	6.5	15.6	67.7	0.9	0.1	0.9	1.2	1.7	5.1
DISP	0.3	9.5	3.1	77.4	1.0	0.1	0.2	1.5	4.4	2.5
FIND	0.3	11.0	1.8	5.5	67.9	0.0	1.4	1.0	3.7	2.5
RETR	0.0	33.3	5.6	38.9	0.0	16.7	0.0	5.6	0.0	0.0
MODS	0.0	8.2	0.0	30.9	3.3	0.0	54.5	2.4	0.8	0.0
STAT	1.3	4.7	1.3	56.4	2.0	0.0	10.7	10.1	11.4	2.0
ENDS	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ERRS	2.4	10.6	3.4	13.4	0.5	0.0	0.2	1.0	8.7	59.8

Table 7. Transition Probability Matrix
Primary Mapping - 1st Order - Context Validated - Predecessor - LC
March-June 1982 (Expressed in Percent)

	BGNS	BRWS	SRCH	DISP	FIND	RETR	MODS	STAT	ENDS	ERRS
BGNS	15.5	19.1	0.4	0.2	1.5	5.6	0.9	1.3	12.5	5.8
BRWS	3.3	48.3	65.2	2.2	23.2	50.0	1.6	10.1	21.5	10.0
SRCH	0.6	2.8	15.6	14.5	1.0	5.6	8.1	9.4	3.2	6.5
DISP	2.2	19.2	14.2	77.4	5.3	22.2	8.9	53.7	40.0	14.9
FIND	0.4	6.1	1.6	1.0	67.8	0.0	11.4	6.7	6.4	2.7
RETR	0.0	0.2	0.1	0.1	0.0	16.7	0.0	0.7	0.0	0.0
MODS	0.0	0.4	0.0	0.7	0.4	0.0	54.5	2.0	0.2	0.0
STAT	0.3	0.3	0.2	1.6	0.3	0.0	13.0	10.1	2.9	0.3
ENDS	75.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ERRS	2.8	3.7	2.7	2.3	0.5	0.0	1.6	6.0	13.3	59.8

Table 8. Transition Probability Matrix
Primary Mapping - 1st Order - Context Validated - Successor - LC
March 1982 (Expressed in Percent)

	BGNS	BRWS	SRCH	DISP	FIND	RETR	MODS	STAT	ENDS	ERRS
BGNS	18.8	61.0	0.8	1.0	1.3	0.2	0.2	0.2	9.2	7.3
BRWS	1.1	51.4	23.2	4.7	11.5	0.1	0.1	0.8	3.6	3.8
SRCH	0.6	8.5	15.6	64.9	1.1	0.0	0.9	2.0	1.2	5.1
DISP	0.3	7.9	2.6	78.5	1.1	0.0	0.3	2.0	4.8	2.7
FIND	0.1	19.8	1.3	5.5	64.2	0.0	1.7	0.9	3.5	2.8
RETR	0.0	0.0	0.0	33.3	0.0	33.3	0.0	33.3	0.0	0.0
MODS	0.0	9.5	0.0	27.6	3.8	0.0	56.2	2.9	0.0	0.0
STAT	0.8	3.2	1.6	58.1	2.4	0.0	11.3	8.9	11.3	2.4
ENDS	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ERRS	2.3	9.8	3.4	12.3	0.4	0.0	0.3	1.0	7.0	63.4

Table 9. Transition Probability Matrix
Primary Mapping - 1st Order - Context Validated - Successor - LC
June 1982 (Expressed in Percent)

	BGNS	BRWS	SRCH	DISP	FIND	RETR	MODS	STAT	ENDS	ERRS
BGNS	9.0	70.5	0.4	1.5	3.0	0.0	0.0	0.4	9.7	5.6
BRWS	0.8	41.7	39.1	3.7	3.4	0.9	0.1	0.1	7.4	2.7
SRCH	0.2	4.0	15.4	71.3	0.6	0.2	0.8	0.2	2.2	5.1
DISP	0.4	12.2	3.9	75.6	0.9	0.2	0.1	0.7	3.8	2.3
FIND	0.7	5.5	3.0	5.5	77.9	0.0	0.4	1.1	4.4	1.5
RETR	0.0	40.0	6.7	40.0	14.4	0.0	0.0	0.0	0.0	0.0
MODS	0.0	0.0	0.0	50.0	0.0	0.0	44.4	0.0	55.6	0.0
STAT	4.0	12.0	0.0	48.0	0.0	0.0	8.0	16.0	12.0	0.0
ENDS	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ERRS	2.8	13.4	3.2	16.7	0.9	0.0	0.0	0.9	13.9	48.1

Table 10. Transition Probability Matrix
Primary Mapping - 2nd Order - Context Validated - LC
March-June 1982 (Expressed in Percent)

	BGNS	BRWS	SRCH	DISP	FIND	RETR	MODS	STAT	ENDS	ERRS
BGNS	8.5	28.5	30.2	5.2	7.3	1.0	0.4	0.1	9.1	9.7
BRWS	1.3	40.3	11.4	29.3	6.2	0.1	0.8	1.0	4.2	5.4
SRCH	0.6	23.7	18.4	41.5	1.7	0.1	0.9	1.2	5.5	6.5
DISP	0.5	9.1	6.7	73.3	1.5	0.1	0.4	1.2	4.4	3.0
FIND	0.1	12.1	1.8	9.0	68.6	0.0	1.2	0.6	3.8	2.7
RETR	0.0	33.4	11.1	38.9	0.0	11.1	0.0	5.6	0.0	0.0
MODS	0.0	9.0	2.5	36.1	4.1	0.0	43.4	3.3	1.6	0.0
STAT	1.5	5.3	2.3	51.5	1.5	0.0	3.0	22.0	7.6	5.3
ENDS	6.9	77.8	0.7	1.4	2.2	0.2	0.2	0.2	3.9	6.9
ERRS	3.2	11.4	8.2	17.0	2.3	0.1	0.2	0.8	7.4	49.3

Table 11. Transition Probability Matrix
Primary Mapping - 3rd Order - Context Validated - LC
March-June 1982 (Expressed in Percent)

	BGNS	BRWS	SRCH	DISP	FIND	RETR	MODS	STAT	ENDS	ERRS
BGNS	7.2	22.5	11.5	33.5	6.6	0.2	1.1	0.8	7.2	9.4
BRWS	1.2	40.5	10.2	27.0	8.1	0.1	0.8	0.9	5.5	3.7
SRCH	0.7	19.5	23.4	40.5	2.2	0.1	0.7	1.7	5.1	6.4
DISP	0.6	9.0	5.9	73.8	1.7	0.1	0.3	1.5	3.9	3.1
FIND	0.6	14.1	2.2	11.8	62.0	0.0	1.6	0.9	3.8	3.0
RETR	5.6	0.0	27.2	44.4	0.0	11.0	5.6	0.0	11.1	0.0
MODS	0.0	10.0	0.8	39.2	4.2	0.0	38.3	4.2	3.3	0.0
STAT	0.0	10.7	2.5	51.1	0.8	0.0	3.3	9.8	10.7	8.2
ENDS	1.8	31.0	34.9	5.3	8.3	1.1	0.5	0.2	7.9	9.2
ERRS	3.8	10.4	7.5	27.3	2.5	0.1	0.4	0.8	5.4	47.0

Table 12. Transition Probability Matrix from State J to K
Primary Mapping - 4th Order - Context Validated - LC
March-June 1982 (Expressed in Percent)

	BGNS	BRWS	SRCH	DISP	FIND	RETR	MODS	STAT	ENDS	ERRS
BGNS	6.8	21.0	8.5	34.7	8.0	0.2	0.3	0.8	10.1	9.8
BRWS	1.3	37.5	13.4	27.5	7.8	0.1	0.8	1.1	4.9	5.6
SRCH	1.0	17.2	17.8	47.3	2.5	0.2	0.5	1.5	5.4	6.6
DISP	0.6	10.7	5.8	72.2	2.0	0.1	0.4	1.3	3.6	3.4
FIND	0.6	13.5	2.0	14.0	60.0	0.0	1.7	0.8	4.7	3.0
RETR	0.0	12.5	0.0	56.3	0.0	6.3	6.3	0.0	12.5	6.3
MODS	0.0	10.3	1.7	40.5	5.2	0.0	37.1	2.6	1.7	1.0
STAT	0.0	6.4	6.4	50.5	1.8	0.0	1.8	18.3	9.2	5.5
ENDS	1.9	23.1	12.2	37.5	7.1	0.2	1.3	1.0	7.1	8.6
ERRS	3.3	11.9	6.4	24.0	3.0	0.0	0.8	1.0	4.4	45.2

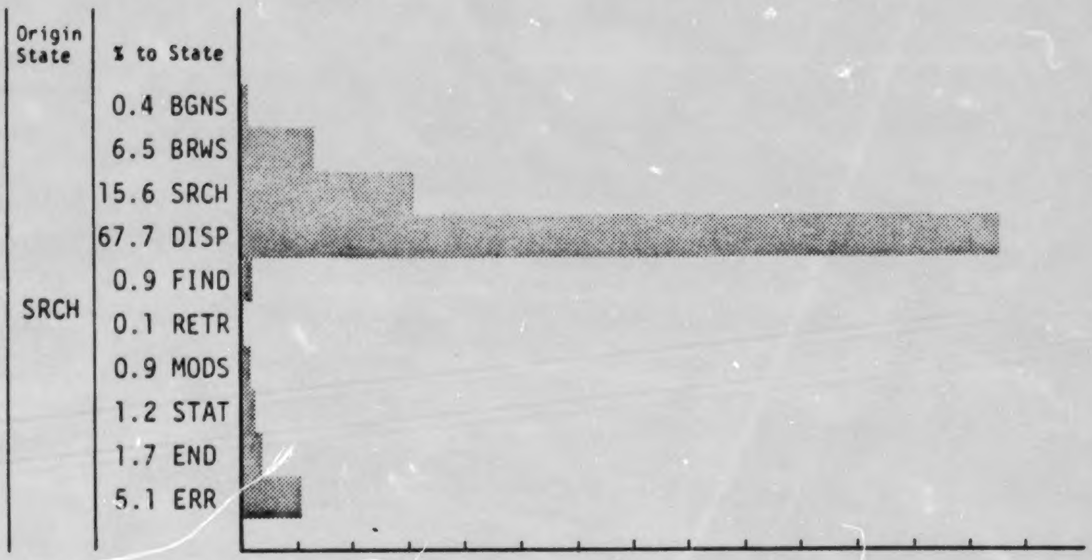


Figure 4. State Transition Diagram from SRCH State
Primary Mapping - 1st Order - LC - March-June 1982

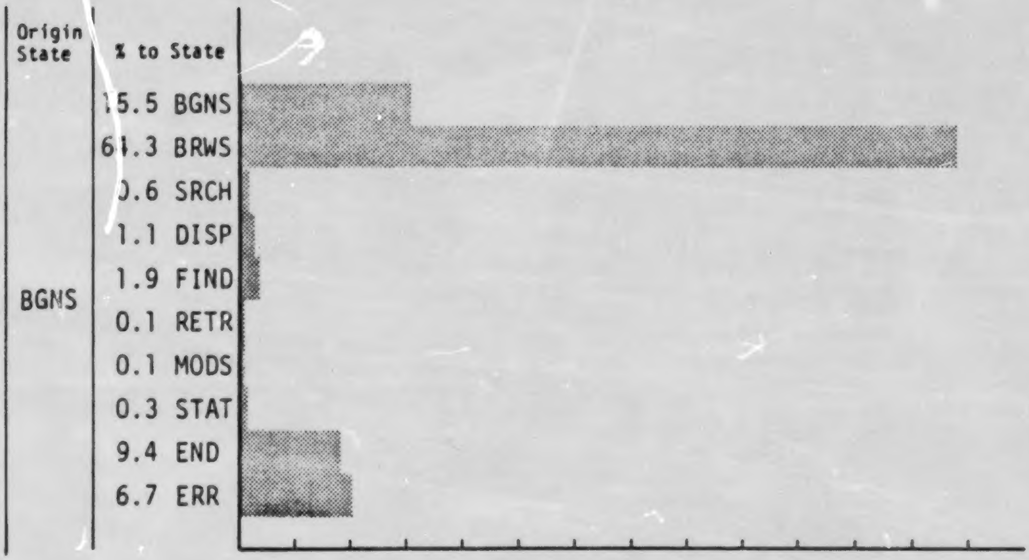


Figure 5. State Transition Diagram from BGNS State
Primary Mapping - 1st Order - LC - March-June 1982

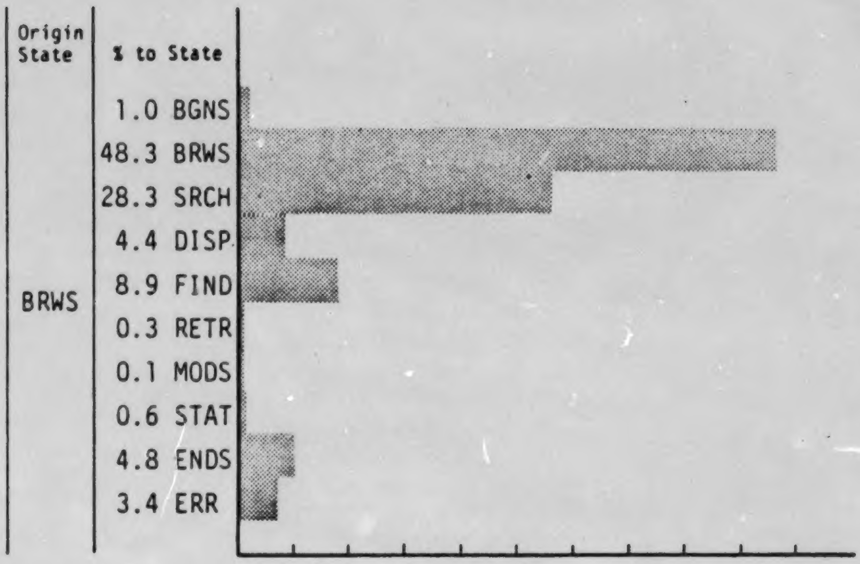


Figure 6. State Transition Diagram from BRWS State
Primary Mapping - 1st Order - LC - March-June 1982

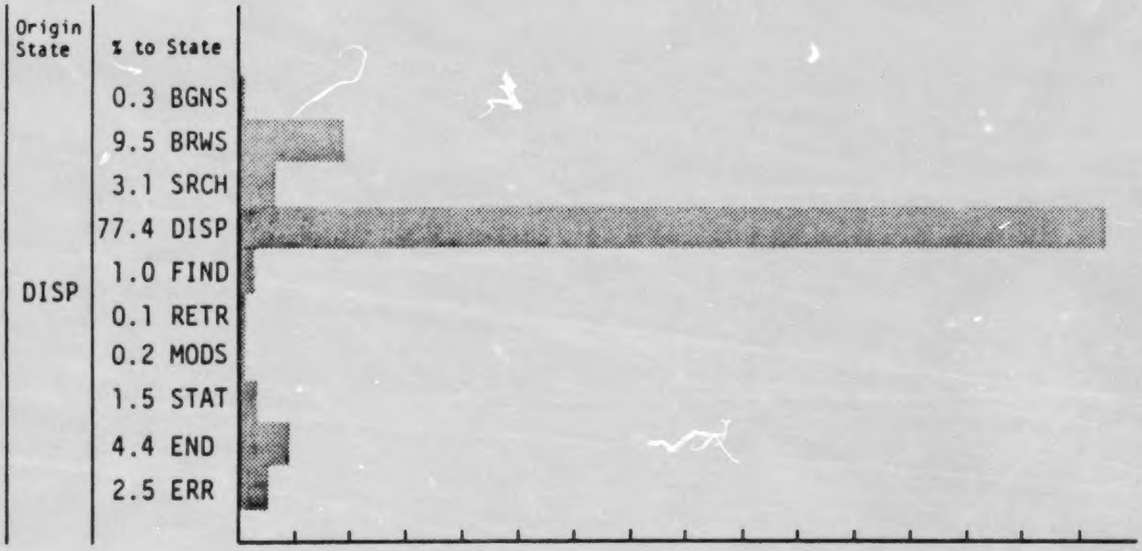


Figure 7. State Transition Diagram from DISP State
Primary Mapping - 1st Order - LC - March-June 1982

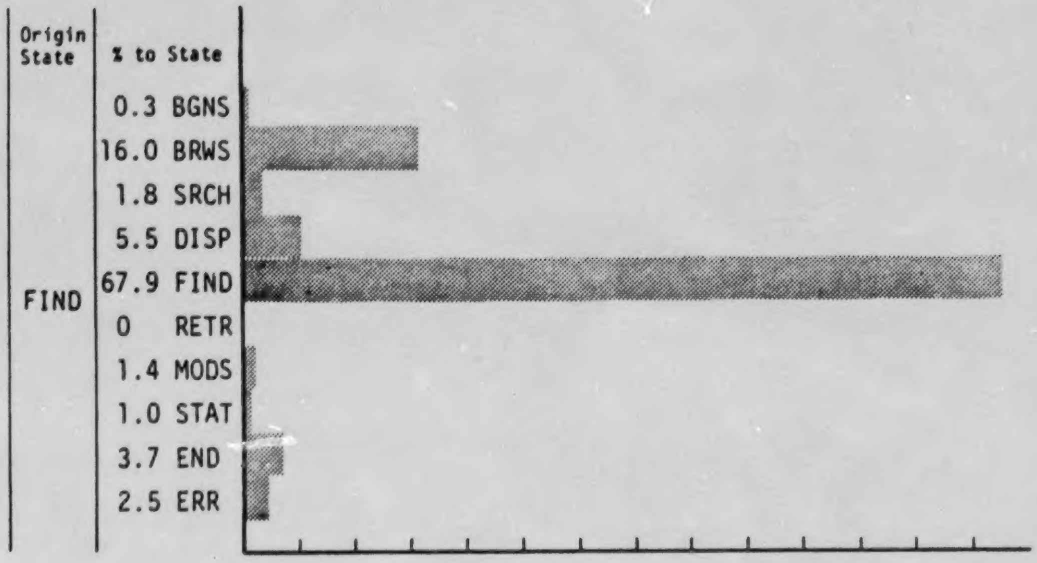


Figure 8. State Transition Diagram from FIND State
Primary Mapping - 1st Order - LC - March-June 1982

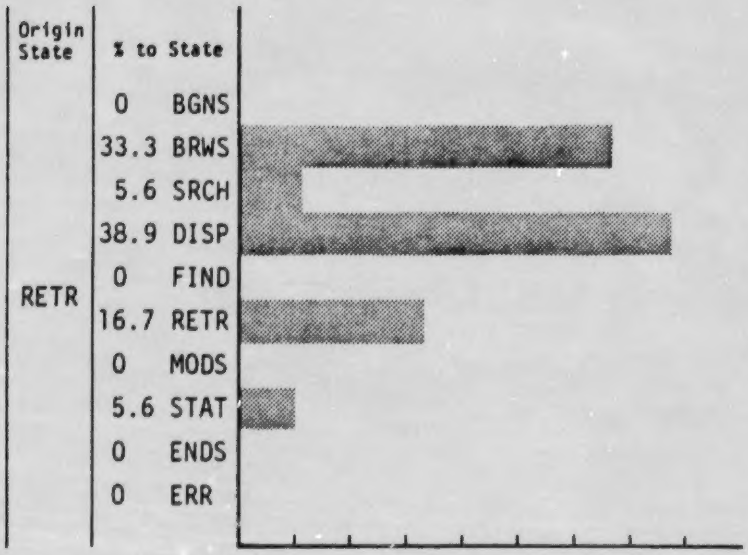


Figure 9. State Transition Diagram from RETR State
Primary Mapping - 1st Order - LC - March-June 1982

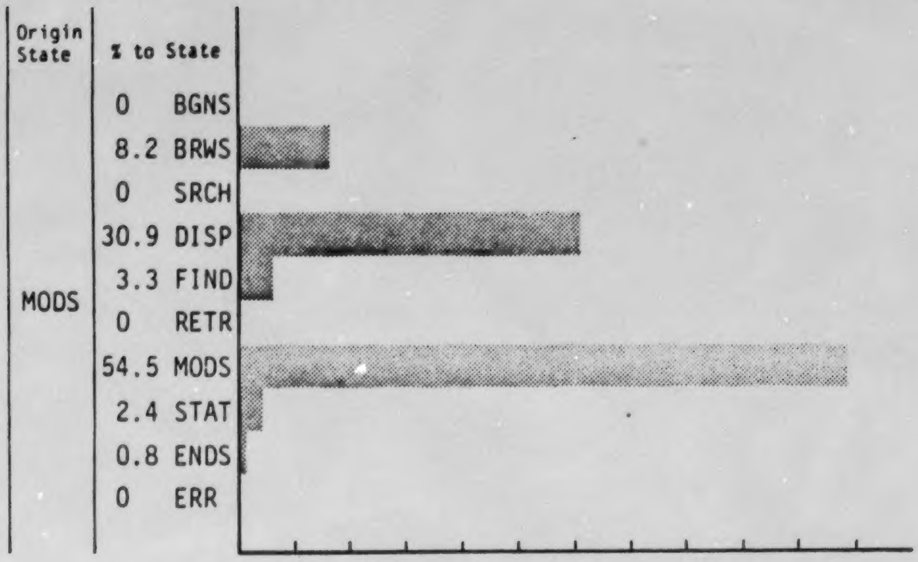


Figure 10. State Transition Diagram from MODS State
Primary Mapping - 1st Order - LC - March-June 1982

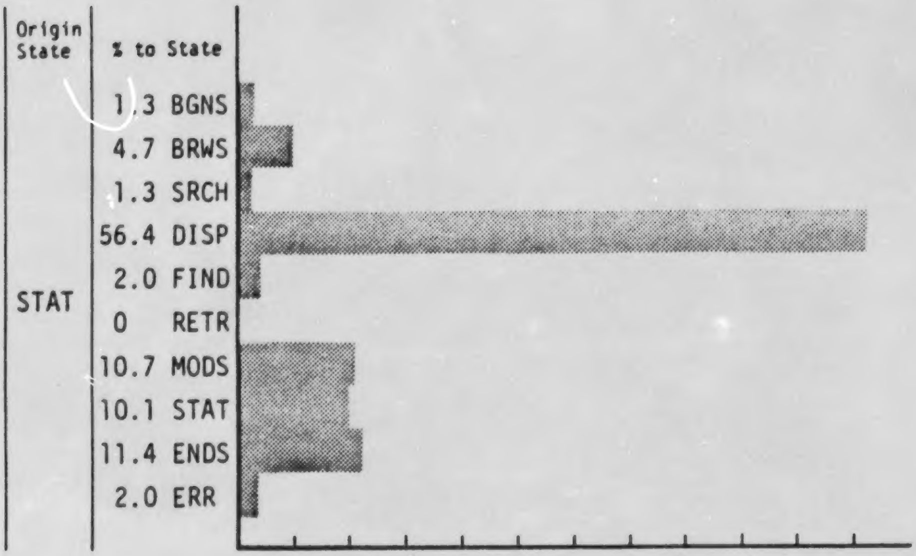


Figure 11. State Transition Diagram from STAT State
Primary Mapping - 1st Order - LC - March-June 1982

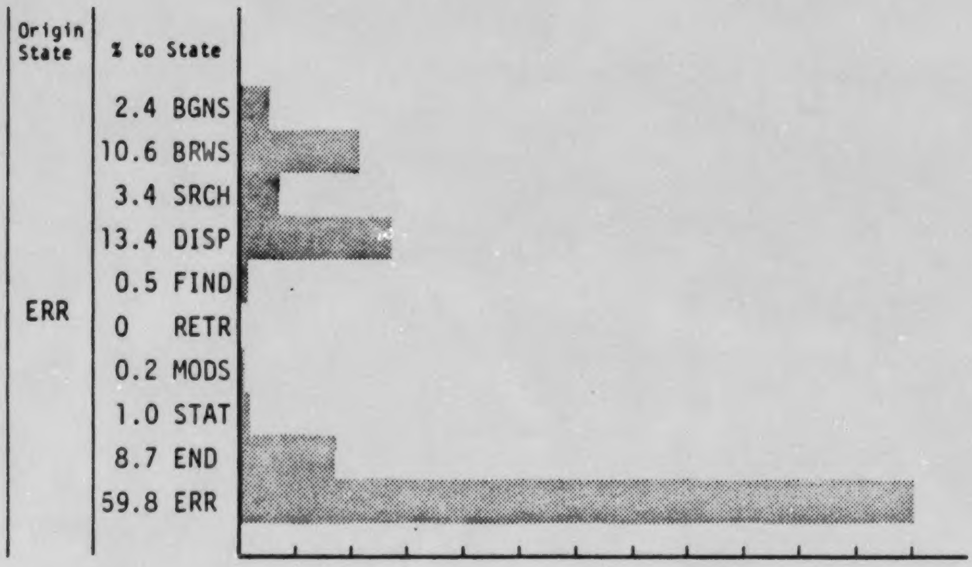


Figure 12. State Transition Diagram from ERR State
Primary Mapping - 1st Order - LC - March-June 1982

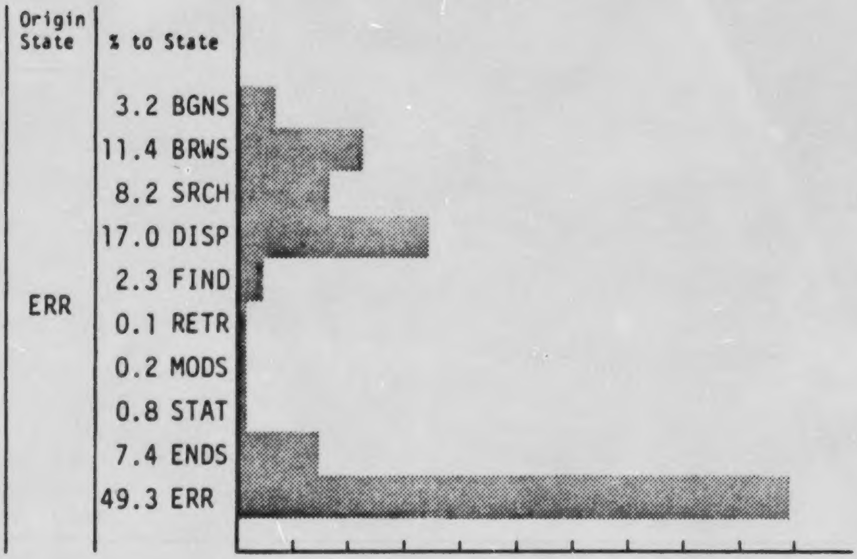


Figure 13. State Transition Diagram from ERR State
Primary Mapping - 2nd Order - LC - March-June 1982

4.3 SYRACUSE UNIVERSITY

The Syracuse system contains one database and one retrieval system for searching it. Essentially, this search system contains search commands, display commands, and help requests. There are six search commands commonly used by public users: author search, title search, subject search, combined subject title word search, call number search, and series title search. In addition, there are several other specialized searches including ones for accession number and UCLC number, but these are rarely used on public terminals.

Two special features are available with the search commands. One allows for the use of partial words. When a person is unsure of the complete spelling of a word, he may enter as much of it as he knows followed by a "+". This feature, however, was not examined in the present analysis.

The other feature allows for the combination of searches. A person may enter two or more searches at a single command. This is equivalent to the Boolean "ENDING" function and will yield the subset corresponding to the interaction or overlap in the returns from the two searches.

There are three display commands that can be entered if hits have been found for a search. When the user receives the "HIT" message, he is informed of the number of hits found. He is directed to press the return key to see some (four to five) of these displayed on the screen. This message will be reported on the screen until the final item in this set of hits is displayed. A person can skip around, however, by asking for the display to begin with any item number. For instance, if the user gets 50 hits, he can ask for the display beginning with the 23rd item (i.e., items 23-27). Finally, a user may ask for more detailed information on any particular item. The HELP request can be entered any time the user becomes confused.

Data was collected at Syracuse University on each of the following dates:

*820412 (Monday)
*820413 (Tuesday)
*820414 (Wednesday)
*820415 (Thursday)
*820418 (Sunday)
*820419 (Monday)
*820420 (Tuesday)
*820421 (Wednesday)
820422 (Thursday)
820423 (Friday)
820424 (Saturday)
820425 (Sunday)
820426 (Monday)
*820427 (Tuesday)
820428 (Wednesday)

Since mislabeled tapes remained undiscovered until late in the analysis, only the data from the dates marked with asterisks could be analyzed in time for this report. These dates represent 62.6% of the total data collected. Unfortunately, only one weekend day (Sunday, April 18) was included in the analysis, and this data apparently represented less than an entire day of transactions (it had only one-sixth of the number of transactions found in the file for Sunday, April 25.) This effectively precluded any analysis comparing weekday and weekend use.

Excluding the data from Sunday (which represents less than 1% of the data analyzed), this sample represents a section of weekdays exhibiting an early-week bias. Therefore, this analysis cannot conclude anything definitive concerning late-week user behavior. Similarly, this data comes from a two-week period in April, and this also limits the ability to generalize the study to this time of year and the end of the term. Finally, since the data from all public use terminals in all buildings at Syracuse University was pooled prior to analysis, we cannot conclude that the results describe behavior at any one location.

A point should be emphasized here about the ability to discern session boundaries when they are not given. In an attempt to estimate session boundaries, various time gaps between commands were considered assuming a relation between gaps and session endings. A null session was defined as a session transitioned from a begin session to an end session. By looking at the content of another user's search, we determined the number of null sessions as well as the number of sessions, and reached the following conclusions:

1. There are a significant number of long time gaps (i.e., those in excess of five minutes). Additionally, one does not start picking up large numbers of new sessions until dropping below a 3 minute criterion, at which point the number of new sessions generated would rise sharply. While by no means certain, it is reasonable to assume that this rise relates to a pick up of type 1 errors (i.e., false boundaries due to intra-session pauses).

2. The vast majority of null sessions can be eliminated by the use of a 3 minute boundary. It is again reasonable to assume that these null sessions are created by intra-session pauses surrounding user activities not targeted in the map used. This could happen when patrons approach the terminal, input an 'S' to indicate a new search, and then hestiate for more than a minute.

3. The vast majority of sessions beginning with a display command are also eliminated when the 3 minute boundary is used.

4. In terms of a graph, all three of these measures display sharp changes at the 3 minute boundary indicating that the decrease in errors as a function of the length of the inter-session gap is slow for longer boundary durations.

5. This type of criterion analysis is aimed at decreasing type 1 errors (false boundaries). Since we do not know how to identify good boundaries, it is not possible to say what the type 2 error rate is (missed boundaries). It may be very high if there are significant numbers of sessions which start sooner than 3 minutes after the last session. This would occur especially in high traffic times or when there is a queueing problem because too few terminals are available to users.

This session determination should be kept in mind when considering session lengths, both in terms of time and number of commands per session. We cannot claim with any certainty to have been able to determine the real session boundaries, although at Ohio State the use of observers enhanced our ability to determine sessions.

Table 13 illustrates the Syracuse University SULIRS (Syracuse University Libraries' Information Retrieval System) commands and the state mapping used (which also considers codes that were generated by SULIRS, since we obtained system response). The following is a listing of system messages and user commands extracted from the SULIRS tapes and coded into a form suitable for analysis.

USER COMMANDS

RTN	A carriage return.
DISP	A carriage return following an items-found message. This command causes one page of postings to be displayed.
S1ST	User input not identifiable as any other command that begins with an "S." When the system is prepared to accept an "S" as a signal for a new search, it will ignore anything following this "S."
SXXX	Same as S1ST.
STRT	A blank followed by an "S." Functions the same as the above two messages.
SOLO	"S" followed by a return. Functions the same as the above three messages.
NULL	Nothing. Used as a placeholder in data file.
DLIN	"D" and some other number used to select a certain item for expanded information during a search.
JUMP	A number input by a user during a search that causes items to be displayed beginning with this item number.
HLP1	A help request that is almost always successful (i.e., gets some help information), but does not exactly specify what help is wanted. It is simply ")HELP" and return.
HLP0	A help request that is often unsuccessful because the user attempts to specify what help is needed. It is ")HELP xxxxx" and return, where "xxxxx" is some command for which help is being requested.
0	Unidentified search code.
1	Author search.
2	Call number search.
3	Subject search.
4	Series title search.
5	Title search.
6	Keyword search.

7-9	Various staff searches (LC #, Accession #, etc.).
11-99	Various dual-entry searches in which more than one search is specified. If there are three searches, only the first two will be indicated, but they are indicated in the order in which they were entered.

SYSTEM INFORMATION

SCH:	Request for a person to enter search commands.	
HT01	Indicates that one item was found and that patron must enter a carriage return to see it or enter "S" and a carriage return to do a new search.	
HT02	Two	items found.
HT03	3-4	items found.
HT04	5-8	items found.
HT05	9-16	items found.
HT06	17-32	items found.
HT07	33-64	items found.
HT08	65-128	items found.
HT09	129-256	items found.
HT10	257-512	items found.
HT11	513-999	items found.
NONE	No items found.	
2COM	Search terms are too common, leading to more than 999 postings.	
CON	Offers opportunity to look at postings found in 2COM search situation. For more than five items found, the HT message repeats until all items have been viewed or until a new search has been initiated.	
END	End of list of postings. At this point, a person may request a new search or enter DLIN.	
NWOR	No words found following search command (i.e., not told what author, title, etc., to search).	
BCOD	Bad or missing search command entered.	

BDIS	Incorrect display command given.
NVAL	An invalid item number during a display request.
XFMT	Incorrect format entered during a call number search command.
SVAL	Invalid system command.
XERR	An unidentified error message (i.e., begins with a "*").

OTHERS

SES1	Greater than 1 minute found between last system message and next user command.
SES2	Greater than 2 minutes found between last system message and next user command.
SES3	Greater than 3 minutes found between last system message and next user command.
SES4	Greater than 4 minutes found between last system message and next user command.
SES5	Greater than 5 minutes found between last system message and next user command.
SESS	Nothing (a placeholder)
UNKN	An unknown or unidentifiable user input.

Table 13. SULIRS - STATES/COMMANDS

PROCESS	STATE MAP	SULIRS CODE
1. Begins	BGNS	Name
2. Create Sets (simple searches)	SEARCH	All search command 0-99
3. Display Records/Citations	DISP	DISP, JUMP, DLIN
4. A Help function with its own command	HELP	HLP1, HLPO
5. ERROR	ERROR	NWOR, BCOD, BDIS, NVAL, XFMT, SVAL, XERR
6. OTHER/Includes system messages	Null	All other commands
7. End session	ENDS	none

The data was formatted and recorded at three stages in the analysis of the information provided by Syracuse University. The transaction log format in Figure 2 shows the form in which each of the separate transactions at the SULIRS system was initially coded for our use. This is different from how the data was provided only in that the original data was received on tapes in an unblocked format. This was broken down into individual records before writing it onto our own tapes as the first stage of the data reduction.

During the next stage of data reduction there were three main goals. The first goal was to try to eliminate the data that was irrelevant to this study. The second goal was to compress the rest of the information into as small an area as possible. The third goal was to introduce markers which could be used in identifying sessions.

The first goal was accomplished by writing an algorithm which eliminated blank lines written by the system, along with irrelevant system messages. Also eliminated were all bibliographic records and information returned to the user by the system. None of the user inputs was eliminated.

The second goal was accomplished in several steps. First, all of the data in the text field which could be recognized was coded. These codes represented both user inputs and system responses. All unrecognized system messages were coded as "UNKN" while unrecognized user inputs were not changed. All codes were four characters long (unknown user inputs were simply truncated to their first four characters). In the next step the date stamps on each user input were systemically overwritten with the value corresponding to the elapsed time between when that input was recorded and when the last system message occurred. This gave a measure of the amount of time necessary to formulate and input the various user commands. The next step toward this second goal was to write the code for the last system message and the present user input together on the user input record (system message first followed by user input in the text fields). Then, where this was possible, the system message record was eliminated. Whenever two codes (system message and user input) appeared on the same line, the 23rd character corresponding to the r/w code was changed to "O." The only time it was not possible to collapse two records into one was when two system messages followed each other (two user inputs never followed each other without some system response intervening). Here a blank field was allowed for the user input.

The third goal was accomplished by writing extra records as markers whenever a time gap of 1, 2, 3, 4 or 5 minutes was found between a system message and a user input. These were written such that whenever there was a gap of more than 5 minutes, the marker for all 5 gaps would appear. When there was a gap over 4 minutes the markers for 1-4 would appear, and so forth. This allowed subsequent transition state analysis using each of these as session markers, and the comparison of the outputs.

The final reduction created a file in which only the system message codes and the user input codes were recorded (eliminating the first 23 characters). This substantially reduced the size of the file. In this file there were 20 codes to a record, and all blank fields were coded "NULL". Also, several other character strings which subsequently had been recognized as irrelevant were also coded "NULL". Finally, all "RTN" codes following system messages declaring that hits had been found were recoded "DISP", and all of these system messages declaring hits were recoded in accordance with the rules outlined in the coding sheet.

4.3.1 Transition Probability Matrices, State Diagrams, and Transaction Statistics

Tables 14 to 30 show state transition matrices calculated for the primary mapping, primary maps with display commands broken out, and primary maps with errors broken out. Figures 14 to 24 show state diagrams corresponding to state transitions. Table 31 shows the frequencies of certain ranges of hits for given search types, such as author, title, subject, etc. Table 32 gives the percent of range of hits given a specific search type. Table 33 reflects the frequency of repetitions of specific search commands, i.e., the number of times a specific command is used consecutively. In 249 sessions, the subject search was used four times in a row. Similarly, an author command was used 2 times in a row a total of 989 times.

All matrices and diagrams are based on the entire set of data used. The primary map refers to the general states described earlier and the first mapping to fit to within this structure. In addition, states were further broken down by differentiating different display commands and different error codes. Different mappings were then made using these alternative states and are referred to as primary mapping with display commands broken out and primary mapping with errors broken out. For the search-only map, only specific search terms are used. The search and search outcome map includes two additional states, DONE and NONE, which represent system messages. DONE indicates that hits were obtained and all of them viewed by the user and NONE indicates no hits were found. Additionally, whenever there is a transition from a search to a search using this map, hits were obtained but not all of the hits were viewed. The "search only normalized" map factors out the repeated specific searches.

Table 14. Transition Frequencies from State A to State B
Primary Map - SULIRS

	BGNS	SRCH	DISP	HELP	ERRS	ENDS	TOTAL
BGNS	0	4327	252	5	1380	446	6409
SRCH	0	13564	18429	26	2950	1347	36316
DISP	0	12376	32744	37	3802	3943	52902
HELP	0	40	28	71	45	20	204
ERRS	0	6009	1449	65	3282	653	11567
ENDS	6409	0	0	0	0	0	6409
TOTAL	6409	36316	52902	204	11459	6409	

Table 15. State Transition Matrix
Percent Starting in State A, Ending in State B
Primary Map - 1st Order - Successor - SULIRS

	BGNS	SRCH	DISP	HELP	ERRS	ENDS
BGNS	0	67.5	3.6	0.1	21.5	7.0
SRCH	0	37.3	50.0	0.1	8.1	3.7
DISP	0	23.4	61.9	0.1	7.2	7.5
HELP	0	19.6	13.2	34.8	22.1	9.8
ERRS	0	52.4	11.6	0.6	28.6	5.7
ENDS	100.0	0	0	0	0	0

Table 16. State Transition Matrix
Percent Starting in State A Given They End in State B
Primary Map - 1st Order - Predecessor - SULIRS

	BGNS	SRCH	DISP	HELP	ERRS	ENDS
BGNS	0	11.9	0.5	2.5	12.0	7.0
SRCH	0	37.3	34.8	18.1	30.9	57.4
DISP	0	34.1	61.9	18.1	33.2	61.5
HELP	0	0.1	0.1	34.8	0.4	0.3
ERRS	0	16.5	2.7	31.9	28.6	10.2
ENDS	100.0	0	0	0	0	0

Table 17. Transition Frequencies from State A to State B
Primary Map with Display Commands Broken Out - 1st Order - SULIRS

	BGNS	SRCH	DISP	JUMP	DLIN	HELP	ERRS	ENDS	TOTAL
BGNS	0	4327	230	10	12	5	1380	446	6409
SRCH	0	13564	18164	71	194	26	2950	1347	36316
DISP	0	11848	31024	218	719	37	3542	3681	51070
JUMP	0	88	124	97	30	0	101	53	493
DLIN	0	440	174	40	318	0	158	209	1339
HELP	0	40	27	1	0	71	45	20	204
ERRS	0	6009	1327	56	66	65	3282	653	11459
ENDS	6409	0	0	0	0	0	0	0	6409
TOTAL	6409	36316	51070	493	1339	204	11459	6409	

Table 18. Percent Transitions Ending in State B Given They Start in State A - Primary Map with Display Commands Broken Out - 1st Order - Successor - SULIRS

	BGNS	SRCH	DISP	JUMP	DLIN	HELP	ERRS	ENDS
BGNS	0	67.5	3.6	0.2	0.2	0.1	21.5	7.0
SRCH	0	37.3	50.0	0.2	0.5	0.1	8.1	3.7
DISP	0	23.2	60.7	0.4	1.4	0.1	6.9	7.2
JUMP	0	17.8	25.2	19.7	6.1	0	20.5	10.8
DLIN	0	32.9	13.0	3.0	23.7	0	11.8	15.6
HELP	0	19.6	13.2	0.5	0	34.8	22.1	9.8
ERRS	0	52.4	11.6	0.5	0.6	0.6	28.6	5.7
ENDS	100.0	0	0	0	0	0	0	0

Table 19. Percent Transitions Starting in State A Given They End in State B - Primary Map with Display Commands Broken Out - 1st Order - Predecessor - SULIRS

	BGNS	SRCH	DISP	JUMP	DLIN	HELP	ERRS	ENDS
BGNS	0	11.9	0.5	2.0	0.9	2.5	12.0	7.0
SRCH	0	37.3	35.6	14.4	14.5	12.7	25.7	21.0
DISP	0	32.6	60.7	44.0	53.7	18.1	30.9	57.4
JUMP	0	0.2	0.2	19.7	2.2	0	0.9	0.8
DLIN	0	1.2	0.3	8.1	23.7	0	1.4	3.3
HELP	0	0.1	0.1	0.2	0	34.8	0.4	0.3
ERRS	0	16.5	2.6	11.4	4.9	31.9	28.6	10.2
ENDS	100.0	0	0	0	0	0	0	0

Table 20. State A - State B Transition Frequencies Using Primary Map with Specific Errors Broken Out - 1st Order - SULIRS

	BGNS	SRCH	DISP	HELP	NWOR	BCOD	BDIS	NVAL	XFMT	SVAL	XERR	ENDS	TOTAL
BGNS	0	4327	252	5	115	822	425	8	4	1	4	446	6409
SRCH	0	13564	18429	26	655	1878	295	6	94	2	20	1347	36316
DISP	0	12376	32744	37	94	1268	2215	205	7	1	12	3943	52902
HELP	0	40	28	71	4	31	4	1	4	1	0	20	204
NWOR	0	771	0	3	12	126	0	1	0	0	1	43	957
BCOD	0	3841	6	41	66	2133	0	0	4	3	12	447	6553
BDIS	0	1194	1352	6	11	251	549	10	3	1	1	131	3509
NVAL	0	88	84	5	0	18	15	10	2	1	1	18	242
XFMT	0	84	1	7	0	15	3	1	2	0	1	6	120
SVAL	0	2	1	1	0	4	0	0	0	5	0	2	15
XERR	0	29	5	2	0	7	3	0	0	0	10	6	62
ENDS	6409	0	0	0	0	0	0	0	0	0	0	0	6409
TOTAL	6409	36316	52902	204	957	6553	3509	242	120	15	62	6409	

Table 21. Percent Transitions Ending in State B Given They Start in State A - Primary Map with Specific Errors Broken Out - 1st Order - Successor - SULIRS

	BGNS	SRCH	DISP	HELP	NWOR	BCOD	BDIS	NVAL	XFMT	SVAL	XERR	ENDS
BGNS	0	67.5	3.6	0.1	1.8	12.8	6.6	0.1	0.1	0	0.1	7.0
SRCH	0	37.3	50.0	0.1	1.8	5.2	0.8	0	0.3	0	0.1	3.7
DISP	0	23.4	61.9	0.1	0.2	2.4	4.2	0.4	0	0	0	7.5
HELP	0	19.6	13.2	34.8	2.0	15.2	2.0	0.5	2.0	0.5	0	9.8
NWOR	0	80.6	0	0.3	1.3	13.2	0	0.1	0	0	0.1	4.5
BCOD	0	58.6	0.1	0.6	1.0	32.5	0	0	0.2	0	0.2	6.8
BDIS	0	34.0	38.5	0.2	0.3	7.2	15.6	0.3	0.1	0	0	3.7
NVAL	0	36.4	34.7	2.1	0	7.4	6.2	4.1	0.8	0.4	0.4	7.4
XFMT	0	70.0	0.8	5.8	0	12.5	2.5	0.8	1.7	0	0.8	5.0
SVAL	0	13.3	6.7	6.7	0	26.7	0	0	0	33.3	0	13.3
XERR	0	46.8	8.1	3.2	0	11.3	4.8	0	0	0	16.1	9.7
ENDS	100.0	0	0	0	0	0	0	0	0	0	0	0

Table 22. Percent Transitions Starting in State A Given They End in State B - Primary Map with Specific Errors Broken Out - 1st Order - Predecessor - SULIRS

	BGNS	SRCH	DISP	HELP	NWOR	BCOD	BDIS	NVAL	XFMT	SVAL	XERR	ENDS
BGNS	0	11.9	0.5	2.5	12.0	12.5	12.1	3.3	3.3	6.7	6.5	7.0
SRCH	0	37.3	34.8	18.1	68.4	28.7	8.4	2.5	78.3	13.3	32.3	57.4
DISP	0	34.1	61.9	18.1	9.8	19.3	63.1	84.7	5.8	6.7	19.4	61.5
HELP	0	0.1	0.1	34.8	0.4	0.5	0.1	0.4	3.3	6.7	0	0.3
NWOR	0	2.1	0	1.5	1.3	1.9	0	0.4	0	0	1.6	0.7
BCOD	0	10.6	0	20.1	6.9	32.5	0	0	3.3	20.0	19.4	7.0
BDIS	0	3.3	2.6	2.9	1.1	3.8	15.6	4.1	2.5	6.7	1.6	2.0
NVAL	0	0.2	0.2	2.5	0	0.3	0.4	4.1	1.7	6.7	1.6	0.3
XFMT	0	0.2	0	3.4	0	0.2	0.1	0.4	1.7	0	1.6	0.1
SVAL	0	0	0	0.5	0	0.1	0	0	0	33.3	0	0
XERR	0	0.1	0	1.0	0	0.1	0.1	0	0	0	16.1	0.1
ENDS	100.0	0	0	0	0	0	0	0	0	0	0	0

Table 23. State A - State B Transition Frequencies Using Search and Search Outcome Map - 1st Order - SULIRS

	BGNS	AUTH	CALL	SUBJ	SRTL	TITL	KEYW	DUAL	NUMB	DONE	NONE	ENDS	TOTAL
BGNS	0	1418	248	1236	56	1329	995	316	44	1	1	7	5651
AUTH	0	926	23	151	8	337	125	152	1	3400	2802	822	8747
CALL	0	23	106	17	1	26	14	3	2	547	412	113	1264
SUBJ	0	175	18	1056	7	145	104	37	1	2416	4907	671	9537
SRTL	0	8	2	5	21	13	7	4	0	85	236	20	401
TITL	0	219	12	98	17	516	80	57	1	2801	2975	751	7527
KEYW	0	133	19	124	6	115	830	29	1	2346	2712	652	6967
DUAL	0	24	4	8	0	14	10	49	1	529	1026	139	1804
NUMB	0	0	0	0	0	1	2	0	5	89	67	15	179
DONE	0	2786	488	2602	112	2421	2110	543	73	0	22	1070	12227
NONE	0	3035	344	4240	173	2610	2690	614	50	13	0	1391	15160
ENDS	5651	0	0	0	0	0	0	0	0	0	0	0	5651
TOTAL	5651	8747	1264	9537	401	7527	6967	1804	179	12227	15160	5651	

Table 24. Percent Transitions Ending in State B Given They Start in State A - Search and Search Outcome Map
1st Order - Successor - SULIRS

	BGNS	AUTH	CALL	SUBJ	SRTL	TITL	KEYW	DUAL	NUMB	DONE	NONE	ENDS
BGNS	0	25.1	4.4	21.9	1.0	23.5	17.6	5.6	0.8	0	0	0.1
AUTH	0	10.6	0.3	1.7	0.1	3.9	1.4	1.7	0	38.9	32.0	9.4
CALL	0	1.8	8.4	1.3	0.1	2.1	1.1	0.2	0.2	43.3	32.6	8.9
SUBJ	0	1.8	0.2	11.1	0.1	1.5	1.1	0.4	0	25.3	51.5	7.0
SRTL	0	2.0	0.5	1.2	5.2	3.2	1.7	1.0	0	21.2	58.9	5.0
TITL	0	2.9	0.2	1.3	0.2	6.9	1.1	0.8	0	37.2	39.5	10.0
KEYW	0	1.9	0.3	1.8	0.1	1.7	11.9	0.5	0	33.7	38.9	9.4
DUAL	0	1.3	0.2	0.4	0	0.8	0.6	2.7	0.1	29.3	56.9	7.7
NUMB	0	0	0	0	0	0.6	1.1	0	2.8	49.7	37.4	8.4
DONE	0	22.8	4.0	21.3	0.9	19.8	17.3	4.4	0.6	0	0.2	8.8
NONE	0	20.0	2.3	28.0	1.1	17.2	17.7	4.1	0.3	0.1	0	9.2
ENDS	100.0	0	0	0	0	0	0	0	0	0	0	0

Table 25. Percent Transitions Starting in State A Given They End in State B - Search and Search Outcome Map
1st Order - Predecessor - SULIRS

	BGNS	AUTH	CALL	SUBJ	SRTL	TITL	KEYW	DUAL	NUMB	DONE	NONE	ENDS
BGNS	0	16.2	19.6	13.0	14.0	17.7	14.3	17.5	24.6	0	0	0.1
AUTH	0	10.6	1.8	1.6	2.0	4.5	1.8	8.4	0.6	27.8	18.5	14.5
CALL	0	0.3	8.4	0.2	0.2	0.3	0.2	0.2	1.1	4.5	2.7	2.0
SUBJ	0	2.0	1.4	11.1	1.7	1.9	1.5	2.1	0.6	19.8	32.4	11.9
SRTL	0	0.1	0.2	0.1	5.2	0.2	0.1	0.2	0	0.7	1.6	0.4
TITL	0	2.5	0.9	1.0	4.2	6.9	1.1	3.2	0.6	22.9	19.6	13.3
KEYW	0	1.5	1.5	1.3	1.5	1.5	11.9	1.6	0.6	19.2	17.9	11.5
DUAL	0	0.3	0.3	0.1	0	0.2	0.1	2.7	0.6	4.3	6.8	2.5
NUMB	0	0	0	0	0	0	0	0	2.8	0.7	0.4	0.3
DONE	0	31.9	38.6	27.3	27.9	32.2	30.3	30.1	40.8	0	0.1	18.9
NONE	0	34.7	27.2	44.5	43.1	34.7	38.6	34.0	27.9	0.1	0	24.6
ENDS	100.0	0	0	0	0	0	0	0	0	0	0	0

Table 26. State A - State B Transition Frequencies Using Search Only Map - 1st Order - SULIRS

	BGNS	AUTH	CALL	SUBJ	SRTL	TITL	KEYW	DUAL	NUMB	ENDS	TOTAL
BGNS	0	1420	248	1236	56	1329	995	316	44	7	5651
AUTH	0	4041	107	721	43	1626	568	346	8	1287	8747
CALL	0	115	588	83	16	133	53	28	11	237	1264
SUBJ	0	729	65	5872	46	712	711	120	2	1280	9537
SRTL	0	45	9	59	117	61	43	7	0	60	401
TITL	0	1497	117	731	70	2986	520	228	20	1358	7527
KEYW	0	556	70	704	43	441	3960	100	7	1086	6967
DUAL	0	334	44	128	10	224	110	658	3	293	1804
NUMB	0	10	16	3	0	15	7	1	84	43	179
ENDS	5651	0	0	0	0	0	0	0	0	0	5651
TOTAL	5651	8747	1264	9537	401	7527	6967	1804	179	5651	

Table 27. Percent Transitions Ending in State B Given They Start in State A - Search Only Map 1st Order - Successor - SULIRS

	BGNS	AUTH	CALL	SUBJ	SRTL	TITL	KEYN	DUAL	NUMB	ENDS
BGNS	0	25.1	4.4	21.9	1.0	23.5	17.6	5.6	0.8	0.1
AUTH	0	46.2	1.2	8.2	0.5	18.6	6.5	4.0	0.1	14.7
CALL	0	9.1	46.5	6.6	1.3	10.5	4.2	2.2	0.9	18.8
SUBJ	0	7.6	0.7	61.6	0.5	7.5	7.5	1.3	0	13.4
SRTL	0	11.2	2.2	14.7	29.2	15.2	10.7	1.7	0	15.0
TITL	0	19.9	1.6	9.7	0.9	39.7	6.9	3.0	0.3	18.0
KEYW	0	8.0	1.0	10.1	0.6	6.3	56.8	1.4	0.1	15.6
DUAL	0	18.5	2.4	7.1	0.6	12.4	6.1	36.5	0.2	16.2
NUMB	0	5.6	8.9	1.7	0	8.4	3.9	0.6	46.9	24.0
ENDS	100.0	0	0	0	0	0	0	0	0	0

Table 28. Percent Transitions Starting in State A Given They End in State B - Search Only Map 1st Order - Predecessor - SULIRS

	BGNS	AUTH	CALL	SUBJ	SRTL	TITL	KEYW	DUAL	NUMB	ENDS
BGNS	0	16.2	19.6	13.0	14.0	17.7	14.3	17.5	24.6	0.1
AUTH	0	46.2	8.5	7.6	10.7	21.6	8.2	19.2	4.5	22.8
CALL	0	1.3	46.5	0.9	4.0	1.8	0.8	1.6	6.1	4.2
SUBJ	0	8.3	5.1	61.6	11.5	9.5	10.2	6.7	1.1	22.7
SRTL	0	0.5	0.7	0.6	29.2	0.8	0.6	0.4	0	1.1
TITL	0	17.1	9.3	7.7	17.5	39.7	7.5	12.6	11.2	24.0
KEYW	0	6.4	5.5	7.4	10.7	5.9	56.8	5.5	3.9	19.2
DUAL	0	3.8	3.5	1.3	2.5	3.0	1.6	36.5	1.7	5.2
NUMB	0	0.1	1.3	0	0	0.2	0.1	0.1	46.9	0.8
ENDS	100.0	0	0	0	0	0	0	0	0	0

Table 29. Percent Transitions Ending in State B Given They Start in State A - Search Only Map Normalized to Eliminate Repeated Same Type Searcn - 1st Order - Successor - SULIRS

	BGNS	AUTH	CALL	SUBJ	SRTL	TITL	KEYW	DUAL	NUMB	ENDS
BGNS	0	25.1	4.4	21.9	1.0	23.5	17.6	5.6	0.8	0.1
AUTH	0	0	2.2	15.2	0.9	34.6	12.1	7.4	0.2	27.3
CALL	0	17.0	0	12.3	2.4	19.6	7.9	4.1	1.7	35.2
SUBJ	0	19.8	1.8	0	1.3	19.5	19.5	3.4	0	34.9
SRTL	0	15.8	3.1	20.8	0	21.5	15.1	2.4	0	21.2
TITL	0	33.0	2.7	16.1	1.5	0	11.4	5.0	0.5	29.8
KEYW	0	18.5	2.3	23.4	1.4	14.6	0	3.2	0.2	36.1
DUAL	0	29.1	3.8	11.2	0.9	19.5	9.6	0	0.3	25.5
NUMB	0	10.6	16.8	3.2	0	15.8	7.3	1.1	0	45.2
ENDS	100.0	0	0	0	0	0	0	0	0	0

Table 30. Percent Transitions Starting in State A Given They End in State B - Search Only Map Normalized to Eliminate Repeated Same Type Search - 1st Order - Predecessor - SULIRS

	BGNS	AUTH	CALL	SUBJ	SRTL	TITL	KEYW	DUAL	NUMB	ENDS
BGNS	0	30.1	36.6	33.8	19.8	29.3	33.1	27.5	46.4	0.1
AUTH	0	0	15.9	19.8	15.1	35.8	19.0	30.2	8.5	22.8
CALL	0	2.4	0	2.3	5.6	3.0	1.9	2.5	11.5	4.2
SUBJ	0	15.4	9.5	0	16.2	15.7	23.6	10.5	2.1	22.7
SRTL	0	0.9	1.3	1.6	0	1.3	1.4	0.6	0	1.1
TITL	0	31.8	17.4	20.0	24.7	0	17.4	19.8	21.1	24.0
KEYW	0	11.9	10.3	19.3	15.1	9.8	0	8.7	7.3	19.2
DUAL	0	7.1	6.5	3.4	3.5	5.0	3.7	0	3.2	5.2
NUMB	0	0.2	2.4	0	0	0.3	0.2	0.2	0	0.8
ENDS	100.0	0	0	0	0	0	0	0	0	0

Table 31. Frequencies of Certain Ranges of Hits Given Specific Search Types - SULIRS

Number of Hits	AUTH	CALL	SUBJ	SRTL	TITL	KEYW	DUAL	NUMB
0	2778	409	4374	230	2945	2681	1020	66
1	808	399	575	29	1230	530	396	96
2	577	116	335	12	517	334	111	1
3-4	671	68	400	21	591	475	81	2
5-8	815	45	511	14	528	484	54	3
9-16	869	27	506	15	434	535	29	1
17-32	684	25	513	8	335	428	20	0
33-64	523	16	461	10	262	355	12	1
65-128	322	18	369	15	161	347	5	0
129-256	231	10	270	13	116	187	3	0
257-512	127	7	208	7	76	147	3	0
513->	100	5	213	12	84	206	0	0
	8505	1145	9235	386	7279	6709	1734	170

Table 32. Percent of Range of Hits Given Specific Search Types - SULIRS

Number of Hits	AUTH	CALL	SUBJ	SKTL	TITL	KEYW	DUAL	NUMB
0	32.7	35.7	52.8	59.6	40.5	40.0	58.8	38.8
1	9.5	34.8	6.2	7.5	16.9	7.9	22.8	56.5
2	6.8	10.1	3.6	3.1	7.1	5.0	6.4	0.6
3-4	7.9	5.9	4.3	5.4	8.1	7.1	4.7	1.2
5-8	9.6	3.9	5.5	3.6	7.3	7.2	3.1	1.8
9-16	10.2	2.4	5.5	3.9	6.0	8.0	1.7	0.6
17-32	8.0	2.2	5.6	2.1	4.6	6.4	1.2	0.0
33-64	6.1	1.4	5.0	2.6	3.6	5.3	0.7	0.6
65-128	3.8	1.6	4.0	3.9	2.2	5.2	0.3	0.0
129-256	2.7	0.9	2.9	3.4	1.6	2.8	0.2	0.0
257-512	1.5	0.6	2.3	1.8	1.0	2.2	0.2	0.0
513->	1.2	0.4	2.3	3.1	1.2	3.1	0.0	0.0

Table 33. Frequency of Repetitions of Specific Search Commands - SULIRS

NUMBER OF TIMES COMMAND IS USED CONSECUTIVELY	AUTH	CALL	SUBJ	SRTL	TITL	KEYN	DUAL	NUMB
1	2839	432	1738	213	3015	1589	836	50
2	989	127	733	47	885	617	189	28
3	402	58	412	11	347	316	61	9
4	204	27	249	8	143	159	24	6
5	112	9	170	2	64	105	14	2
6	55	3	101	2	22	63	5	2
7	31	4	60	1	19	38	5	
8	30	4	40		18	35	2	
9	10	3	53		5	25	1	
10	6	2	28		7	14	0	
11	10	1	23		5	13	5	
12	5	0	18		0	10	2	
13	3	1	7		1	6	0	
14	2	1	7		1	2	0	
15	0	1	6		1	6	1	
16	2	1	6		0	1	0	
17	4	0	5		2	2	1	
18	2	0	0		1	1	0	
19	1	0	0		2	0	1	
20	0	0	2		0	2	0	
21	1	0	2		2	0	0	
22		0	0		0	0	0	
23		0	2		0	1	1	
24		0	1		0	1		
25		1	0		0	1		
26		0	0		0	0		
27		0	0		0	0		
28		0	0		1	0		
29		0	0			1		
30		0	0			0		
31		0	0			0		
32		1	0			0		
33			1			0		
34			1			1		

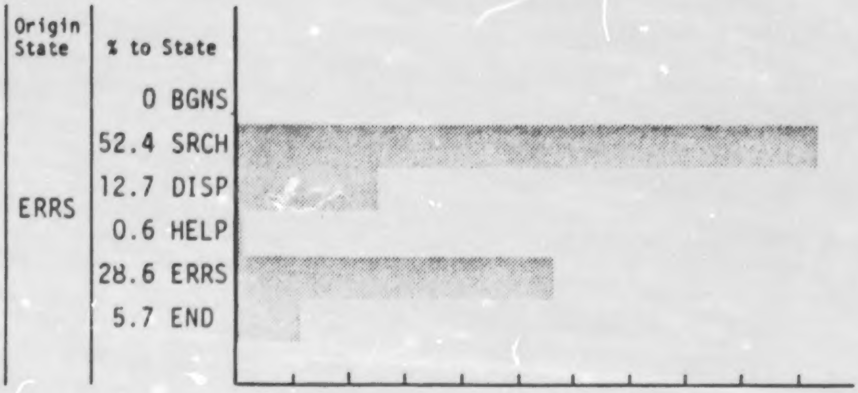


Figure 14. State Transition Diagram from ERRS State
Primary Mapping - 1st Order - SULIRS

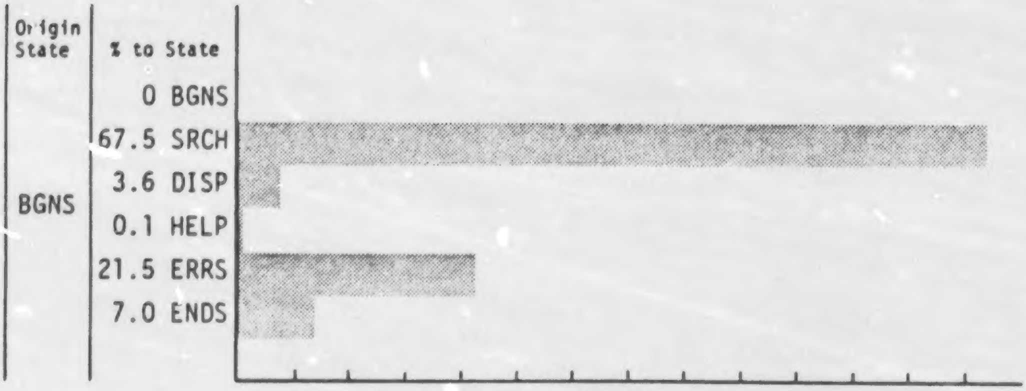


Figure 15. State Transition Diagram from BGNS State
Primary Mapping - 1st Order - SULIRS

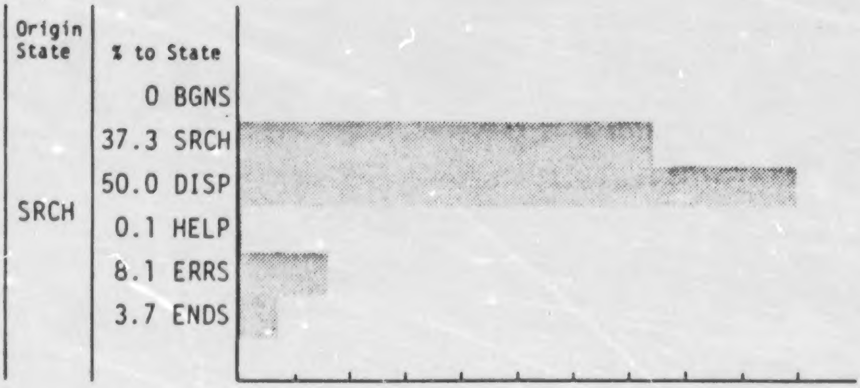


Figure 16. State Transition Diagram from SRCH State
Primary Mapping - 1st Order - SULIRS

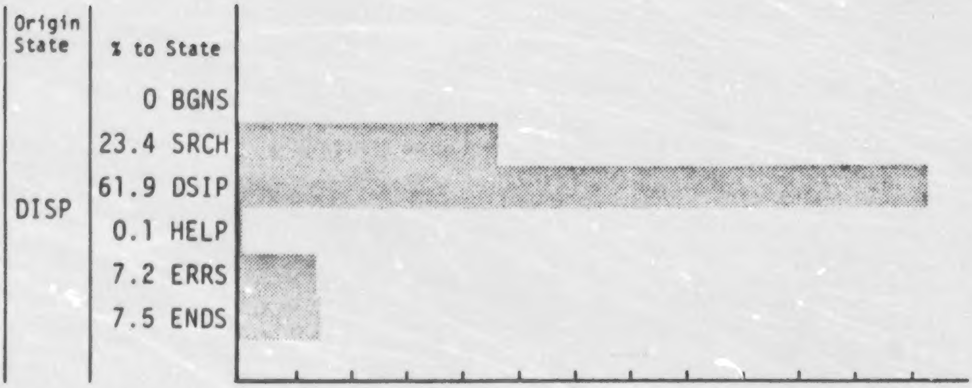


Figure 17. State Transition Diagram from DISP State
Primary Mapping - 1st Order - SULIRS

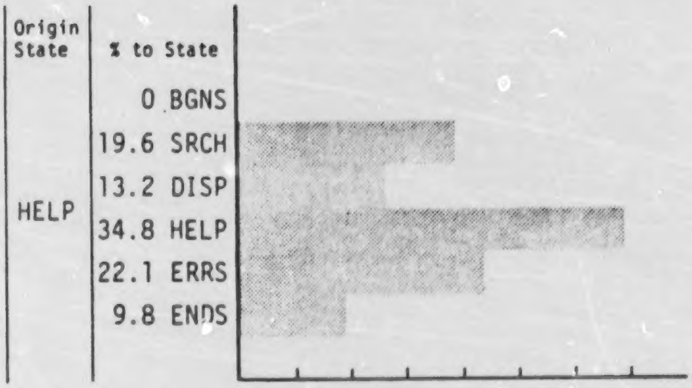


Figure 18. State Transition Diagram from HELP State
Primary Mapping - 1st Order - SULIRS

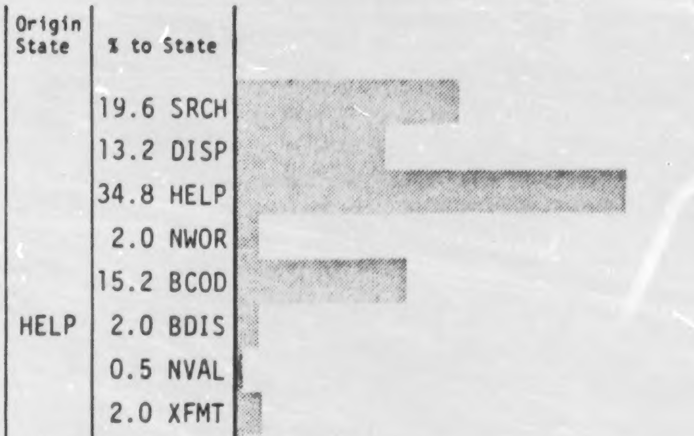


Figure 19. State Transition Diagram from HELP State
Primary Mapping with Explicit Error Commands - 1st Order - SULIRS

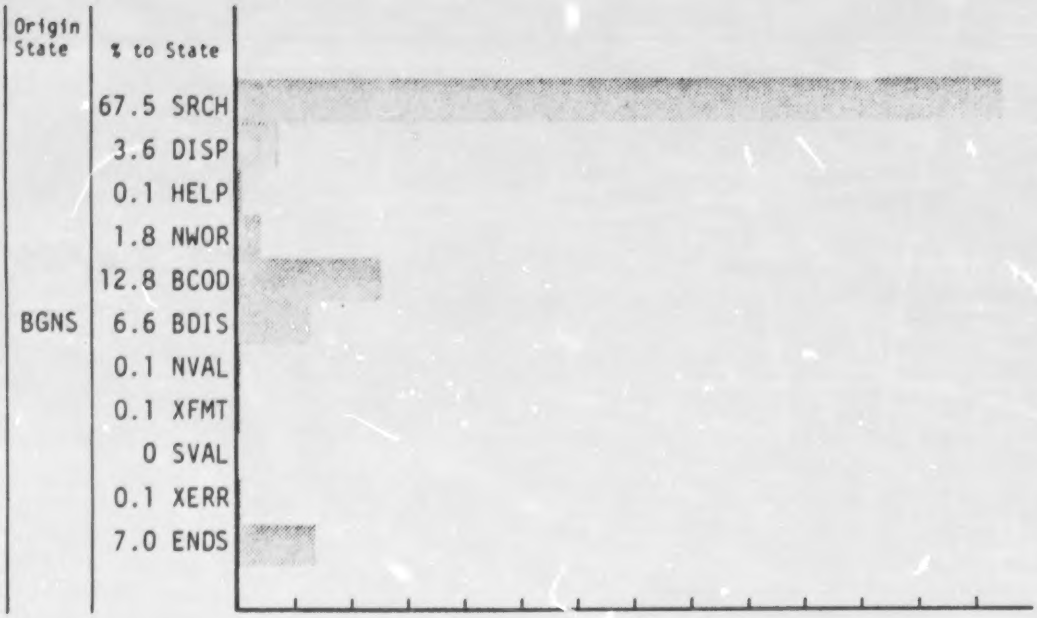


Figure 20. State Transition Diagram from BGNS State
Primary Mapping with Explicit Error Commands - 1st Order - SULIRS

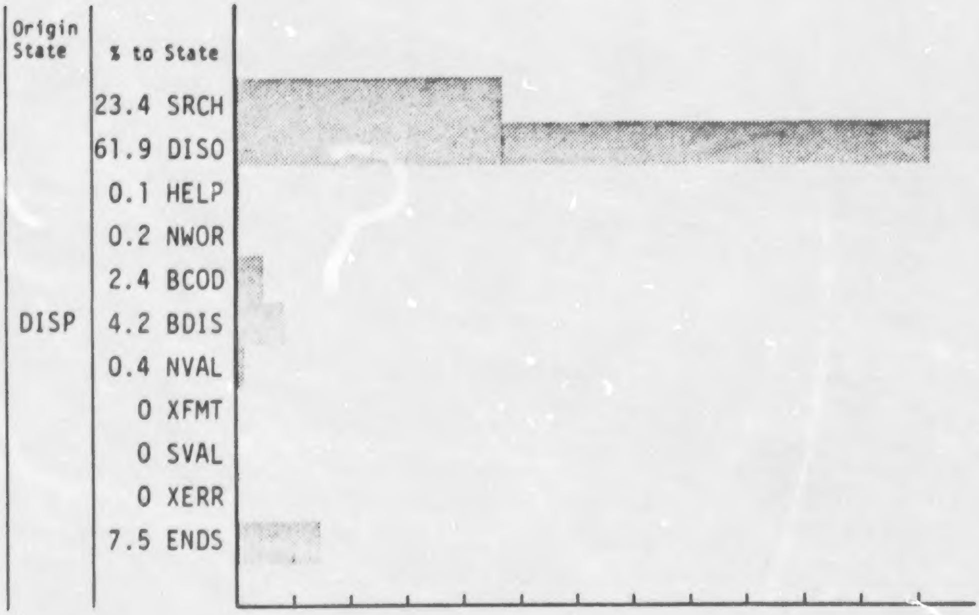


Figure 21. State Transition Diagram from DISP State
Primary Mapping with Explicit Error Commands - 1st Order - SULIRS

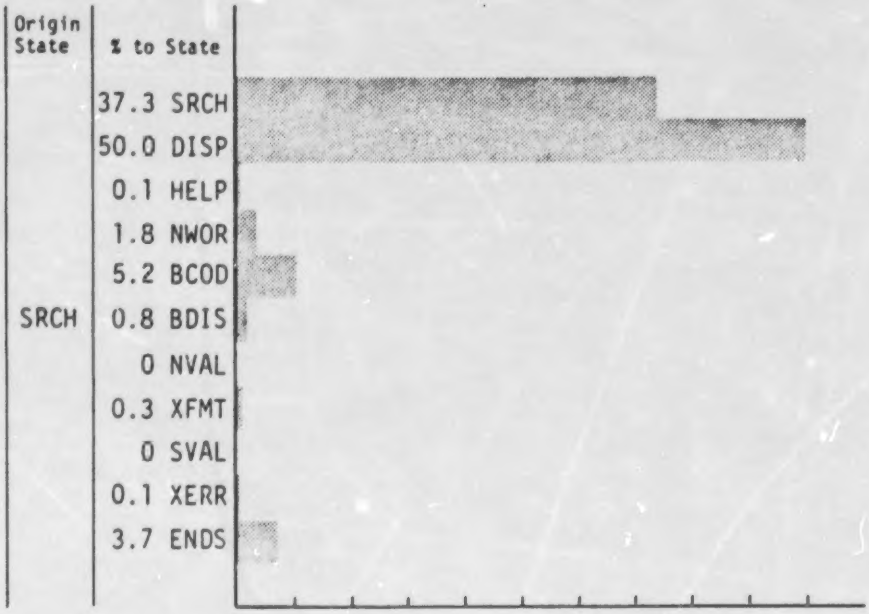


Figure 22. State Transition Diagram from SRCH State
Primary Mapping with Explicit Error Commands - 1st Order - SULIRS

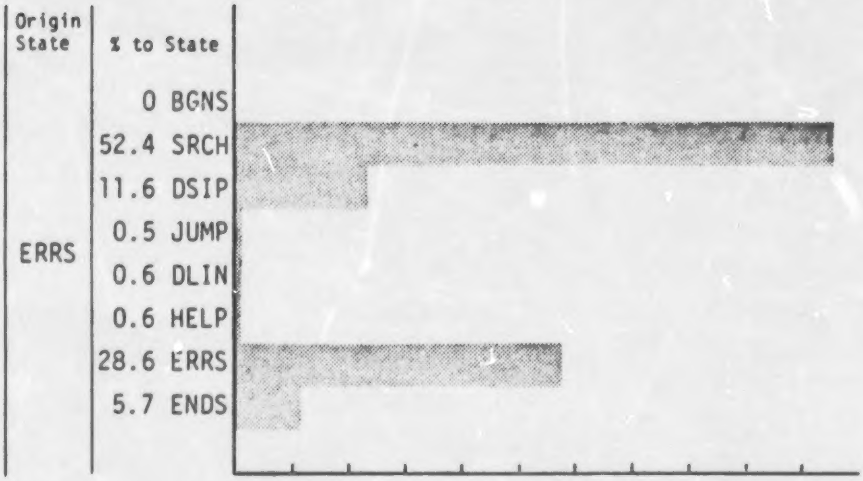


Figure 23. State Transition Diagram from ERRS State
Primary Mapping with Explicit Display Commands - 1st Order - SULIRS

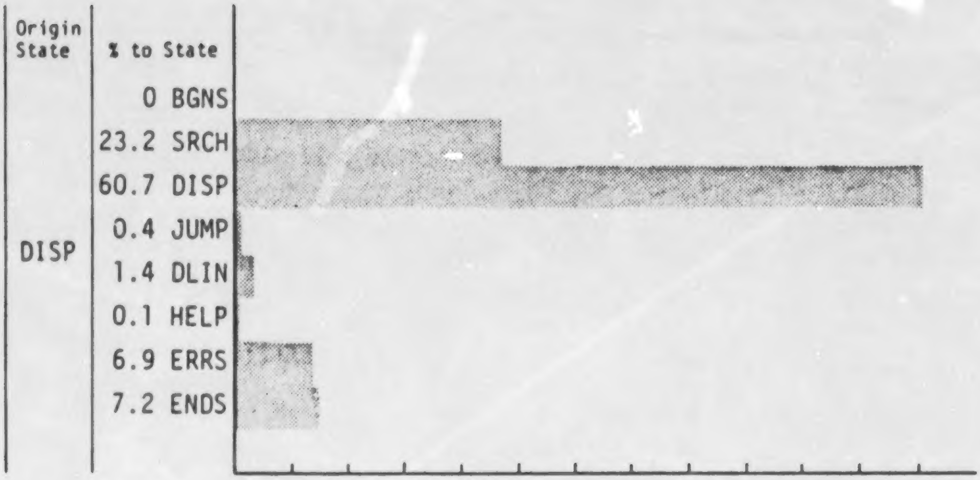


Figure 24. State Transition Diagram from DISP State
Primary Mapping with Explicit Display Commands - 1st Order - SULIRS

4.4 DALLAS PUBLIC LIBRARY

According to the tapes we received, the Dallas Public Library system has 383 terminals. Of these, 73 terminals are for public use, 117 terminals are for staff use, and 193 terminals were used in locations outside the library, such as the Dallas police and fire departments.

The system has four different kinds of terminals. A user can start a search by:

1. Pressing CLEAR key, typing LSCAN, then pressing ENTER key

2. Pressing ERASE PAGE, typing LSCAN, then pressing ENTER

3. Pressing CLEAR key

4. Typing an ampersand in the first position of the subject list field and pressing ENTER. If the LINE # field is present, put the ampersand there (depending on which system he uses).

These procedures did not record on tapes exactly what a user did. Therefore, it was impossible to figure out when a user started and ended a session. After initializing the system, the user's search begins the session.

The system had five different searches available: subject, author, title, author/title, and call number searches. To search, the user tabs the cursor to an appropriate field and types in the content of his search. The user is actually formatting search strings into derived search keys according to the instructions he was given.

After the user types in something to search and obtains a hit, bibliographic information is displayed on the screen. An author search results in 15 items and displayed information includes line #, author name, title, and call number. If the user wants more items, he may press the ENTER key. If something is found, the line # should be entered to get detailed information. After looking at the detailed content, the user may go back to the previous abbreviated display by pressing a key. Detailed information is shown for the line #, ID number, type of material, language, title, DDC number, collection designator, call number, publisher, date of publication, number of material pages, and name of the library where the material is located. One hit gives a full bibliographic record, while more than one hit gives a brief title display.

We received transaction data on tape for the dates August 9 to August 21, 1982. The tapes were 1600 bpi, contained 162,349 records per block, user phase name (always LSCA), date and time in minutes, terminal ID, codes of user activities, and the content of users' searching input. This content can be interpreted as subject search, author search, etc., depending on where it is in the field of a record. Each record of user activities had 86 bytes.

Data was first sorted by user type, user ID, date and time. From the original data, date and time (in minutes), codes, and the content of search input were pulled out and put into 80-byte records. The overall data volume was 334,390 records. The actual content of our record was composed of number of bytes filled in the record, terminal type, terminal ID number, date, time, code, search type, and the content of search.

In the Dallas system, no user commands had to be entered except one that initiated the system for certain terminals. In order to perform a search, the user tabs and fills appropriate fields on the screen. Paging is done by pressing the ENTER key.

To find detailed information for an item on the page, the user types in the line number for that item. If there are more items than can be displayed, and the user wants to see them after the line number, he presses the ENTER key. Should the user want to go back to the previous page after line number select or detailed browsing for a line number select, he presses the PF1 key.

Because there were virtually no commands to be typed in for searches in the Dallas system, the error rates were low.

All statistical analyses were done on the complete set of data obtained from Dallas Public Library for the two-week period. Observation of frequencies from the two different weeks did not reveal significant differences. This finding tended to support the use of one week of data as a representative sample.

Meaningful time analyses on Dallas data were impossible because time stamps were given in minutes.

The Dallas user codes, system codes, and state mappings used in the analysis are as follows:

USER CODES

S1 : Subject search

S2 : Author search

S3 : Title search

S4 : Author/title search

S5 : Call number search

LS : Line number selected

CS : More information for the line number selected

NK : Not known

SYSTEM CODES

A1 : Direct hit

A2 : Search success (displayed as blank, not including direct hit)

MS : Mismatch

ER : Search error (required field has no data)

EA : Search error (non-alphanumeric character not allowed in field)

AL : Line number select succeeded

EL : Line number select error

NK : Not known

CODES WHICH CANNOT BE SEPARATED BY USER AND SYSTEM RESPONSES

PZ : Erase the content on the screen

PA : Page

RT : Retrieving the previous page after LS or CS

NZ : Carriage return (no data entered)

UZ : Data unavailable to computer due to system maintenance work

NL : This code is used to fill two characters to make four characters as a set of user-system responses

00 : Session starts

EE : Session ends

STATE CODES

The state codes used for the first mapping consisted of six states:

1 - Begin

2 - Search

3 - Browse

4 - Display

5 - End

6 - Error

A total of 229,473 transactions were analyzed, and a total of 17,797 sessions were determined by using a maximum time gap of 3 minutes between commands, indicating that a new session began. Although there is a code that may be used to indicate the beginning of a session, it is not required; therefore, a boundary was necessary. An analysis of session differences was made which resulted from using 2-minute, 3-minute, and 5-minute maximum time gaps between commands. This indicated that the 3-minute "session boundary" is not significantly different from a 5-minute boundary in the usage of search commands. In addition, 2-minute session boundaries could produce different results from the 3-minute boundaries. Thus, we decided to play it safe and go with the 3-minute boundary if a new session start was not indicated. This conclusion is supported by the frequency time gaps between "N" codes (which are used for starting a new session) and other data lines in the tapes sent to us. Although this code cannot be used to determine the exact session boundary, it is possible to get an indication of how long a terminal was not in use before a user started to use it again. The frequencies, shown in Table 34, show that if we eliminate time gaps greater than 7 minutes, most of the gaps are less than 3 minutes.

4.4.1 Transition Probability Matrices, State Diagrams, and Statistics

Tables 35 through 37 show the zero through the 2nd order transitions for these mappings. Figures 25 through 34 show the state diagrams. It should be mentioned that our analysis of the data from the two weeks yielded the same patterns; therefore, we used just one week to represent the overall data.

A number of statistics were determined for the Dallas data in addition to the matrices and search patterns. Table 38 summarizes frequencies of 0, 1, 2, 3, 4, and 5 different search types in a session and the frequency of sessions in which there were search types by subject, author, title, author/title, and call number. Table 39 shows the frequency distribution for search types. The frequency of transactions for public and staff users from August 9 to August 21, 1982, by day of week shows very similar activity, except for the activity on Mondays. For example, Table 38 indicates that the number of sessions which contained zero subject searches was 7658, the number with zero author searches was 13561, and zero call number (CAL) was 16318. Also, as examples, there were 1219 occurences of pair searches of subject then title and 480 occurrences of author then title searches. The one search sessions refer to sessions which contained only one search type; 2982 sessions had no search; 9696 sessions had only one kind of search; and 3503 sessions had two different searches.

Table 34. Frequency of Time-Gaps for N Codes - Dallas
(The First Number of Each Row is the Total Frequency of the Following Status of That Row)

time-gaps (in minutes)	total f.	0	1	2	3	4	5	6	7=<
PUBLIC									
Gap between a data line and N code:	28256	12019	7460	1620	867	558	431	360	4941
Gap between N code and N code:	28851	14785	10455	1142	330	224	97	116	1702
Gap between N code and the next line:	6862	3130	202	424	202	112	105	66	1288
STAFF									
Gap between a data line and N code:	8479	1908	1605	619	394	337	216	187	3213
Gap between N code and N code:	9325	5248	2903	325	122	69	45	28	585
Gap between N code and the next line:	1404	399	76	118	76	52	33	22	454

Table 35. 0th Order Matrix - Primary Mapping - Dallas

	FREQUENCY	PERCENTAGE
BEGIN	17797	8
SEARCH	61187	27
BROWSE	97479	42
DISPLAY	30598	13
END	17786	8
ERROR	4626	2
	229473	100

Table 36. First Order Transition Probability Matrix
Primary Mapping - Successor - Dallas

	BGNS	SRCH	BRWS	DISP	ENDS	ERRS
BGNS	----	.82	0	.03	.15	0
SRCH	----	.51	.19	.15	.12	.03
BRWS	----	.09	.70	.18	.04	0
DISP	----	.16	.57	.05	.12	.10
ENDS	----	----	----	----	----	----
ERRS	----	.40	.05	.42	.12	0

Table 37. Second Order Transition Probability Matrix
Primary Mapping - Successor - Dallas

	BGNS	SRCH	BRNS	DISP	END	ERR
BGNS	----	.40	.21	.18	.17	.04
SRCH	----	.45	.34	.10	.12	.02
BRNS	----	.11	.71	.09	.05	.01
DISP	----	.16	.28	.50	.06	0
END	----	----	----	----	----	----
ERR	----	.18	.19	.08	.05	.50

Table 38. Summary of Search Types Within Sessions - Dallas

FREQUENCY OF NO SEARCH IN A SESSION:	2982
FREQUENCY OF ONLY ONE KIND SEARCHES:	9696
FREQUENCY OF ONLY TWO DIFFERENT SEARCHES:	3503
FREQUENCY OF ONLY THREE DIFFERENT SEARCHES:	1246
FREQUENCY OF ONLY FOUR DIFFERENT SEARCHES:	332
FREQUENCY OF ONLY FIVE DIFFERENT SEARCHES:	38

FREQUENCY OF SESSIONS IN WHICH THE FOLLOWING SEARCHES WERE INCLUDED:

SUBJECT	AUTHOR	TITLE	AUT-TITLE	CALL NUMBER
10139	4236	4941	1163	1479

THE ONE-SEARCH-SESSIONS ARE:

SUBJECT	AUTHOR	TITLE	AUT-TITLE	CALL NUMBER
2733	899	1270	223	267

WHICH PAIR SEARCHES IN A SESSION?:

SUB-AUT	SUB-TIT	SUB-ATT	SUB-CAL	AUT-TIT	AUT-ATT	AUT CAL	TIT ATT	TIT CAL	ATT CAL
932	1219	112	382	480	128	68	94	75	13

WHICH TRIPLE SEARCHES IN A SESSION?:

SB-AU-TI	SB-AU-AT	SB-AU-CL	SB-TI-AT	SB-TI-CL	SB-AT-CL	AU-TI-AT	AU-TI-CL	AU-AT-CL	TI-AT-CL
644	97	115	60	158	16	108	24	13	11

TOTAL FREQ.	SUB.	AUT.	TIT.	ATT.	CAL.
0	7658	13561	12856	16634	16318
1	4038	2405	2980	769	909
2	1830	835	981	198	257
3	1121	388	429	83	132
4	761	210	193	41	61
5	549	135	129	29	36
6	372	93	72	11	30
7	308	37	39	6	15
8	220	37	26	8	8
9	158	23	17	4	6
10	109	18	20	3	6
11	131	12	16	1	3
12	91	7	9	1	2
13	78	6	7	1	4
14	57	6	2	0	2
15	41	5	3	1	0
16	28	5	1	1	1
17	38	0	1	1	0
18	26	2	3	0	1
19	30	1	2	1	0
20	15	2	2	1	0
21	22	1	2	1	2
22	12	0	0	1	0
23	18	0	0	0	0
24	13	2	1	0	2
25	7	1	1	0	0
>26	66	5	5	1	2

Table 39. Frequency Distribution for Search Types - Dallas

	FREQUENCY	PERCENTAGE
SUBJECT	36940	37.5
BEGIN	17800	18.1
END	17791	18.1
TITLE	9831	10.0
AUTHOR	9182	9.3
CALL#	3056	3.1
AUT-TIT	2182	2.2
ERROR	1710	1.7
TOTAL	98492	100.0

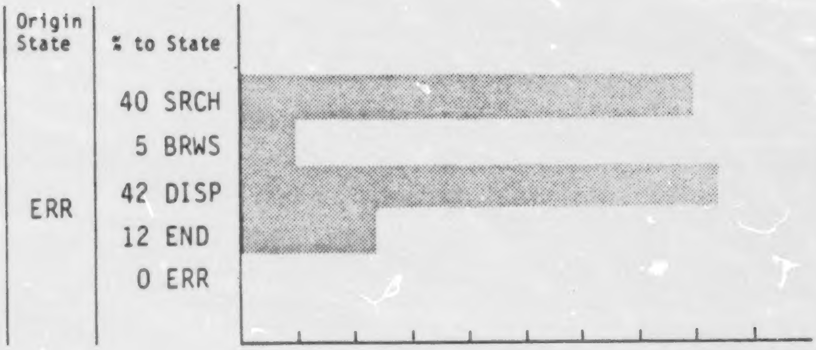


Figure 25. State Transition Diagram
1st Order from ERR - Dallas

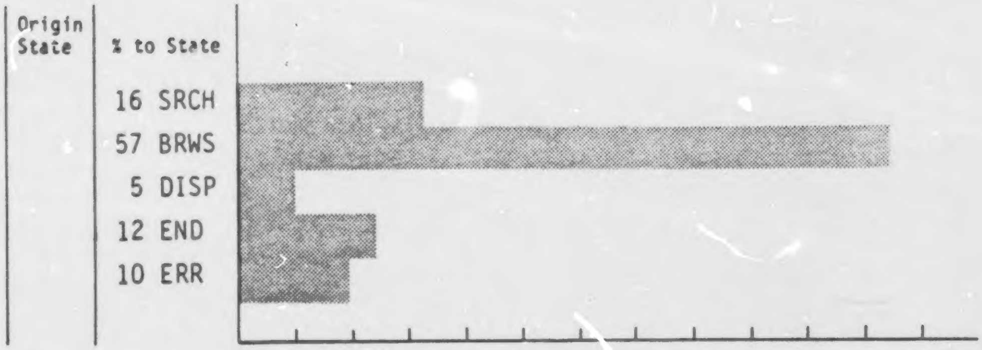


Figure 26. State Transition Diagram
1st Order from DISP - Dallas

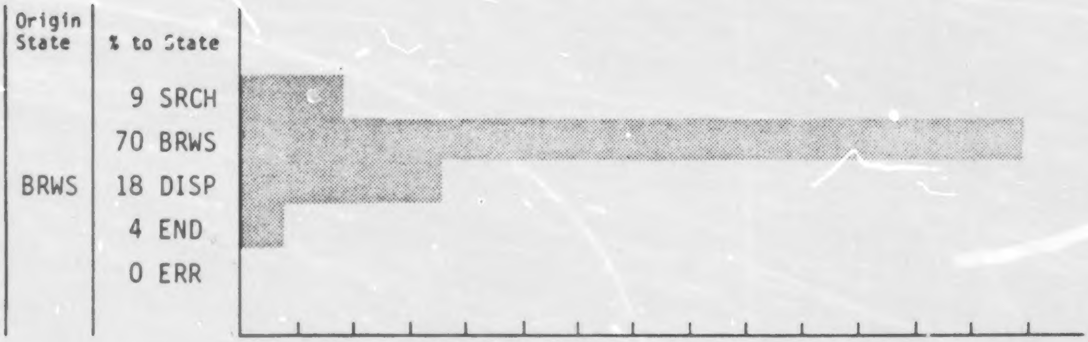


Figure 27. State Transition Diagram
1st Order from BRWS - Dallas

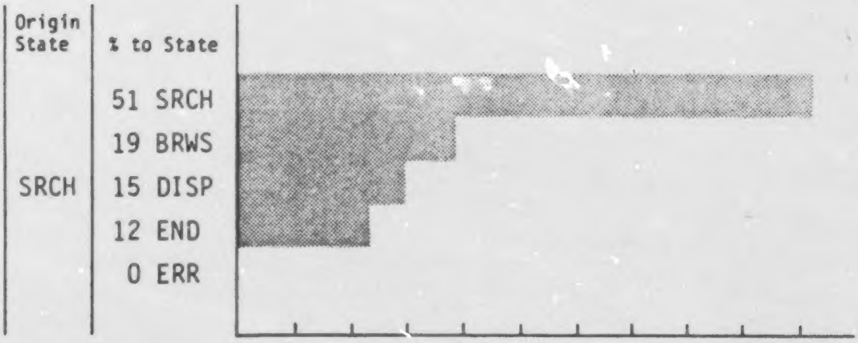


Figure 28. State Transition Diagram
1st Order from SRCH - Dallas

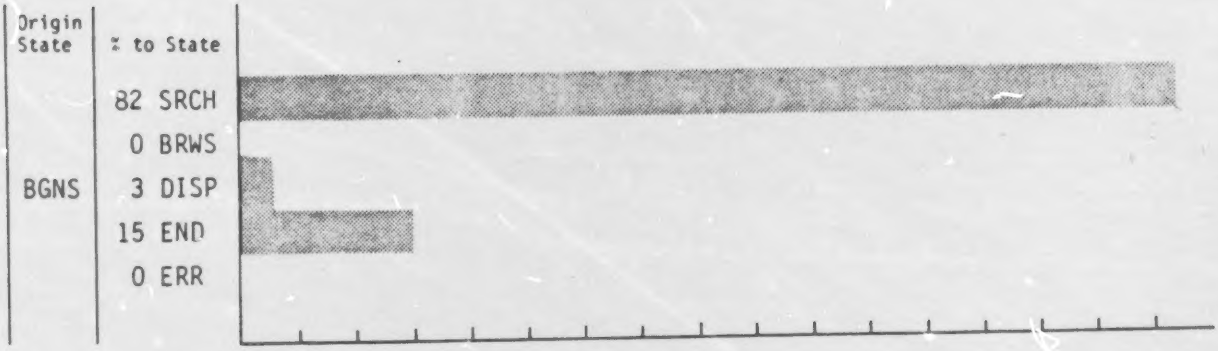


Figure 29. State Transition Diagram
1st Order from BGNS - Dallas

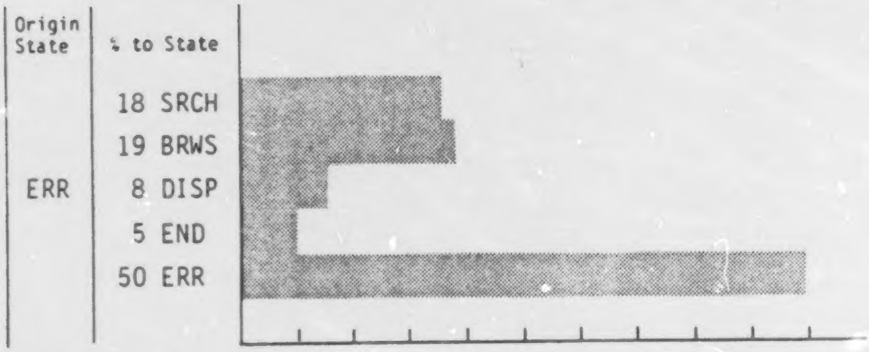


Figure 30. State Transition Diagram
2nd Order from ERR - Dallas

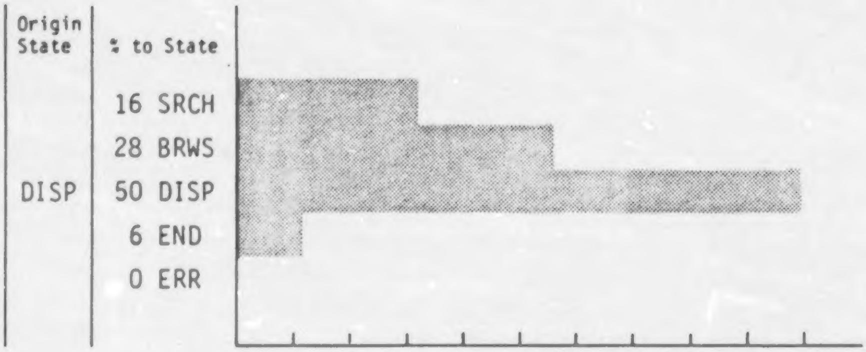


Figure 31. State Transition Diagram
2nd Order from DISP - Dallas

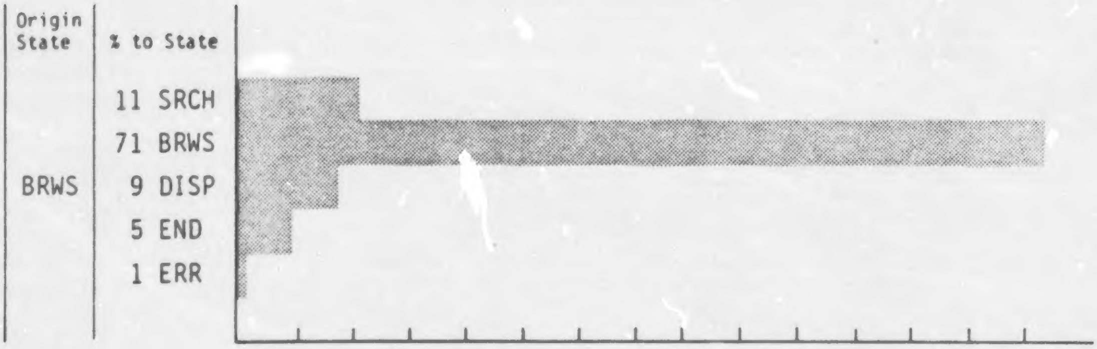


Figure 32. State Transition Diagram
2nd Order from BRWS - Dallas

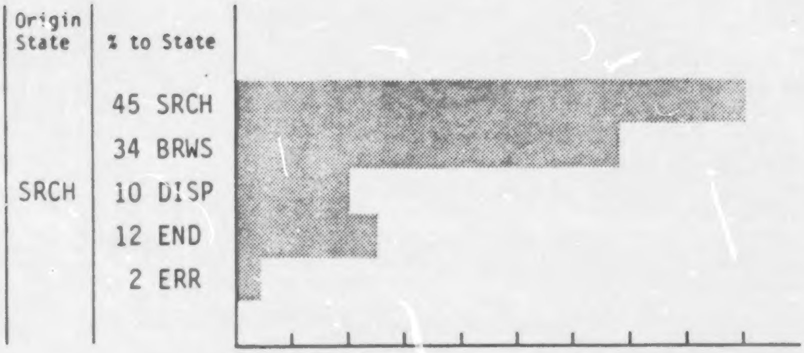


Figure 33. State Transition Diagram
2nd Order from SRCH - Dallas

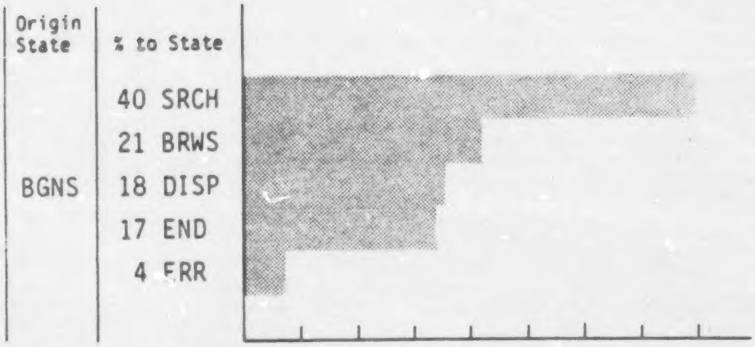


Figure 34. State Transition Diagram
2nd Order from BGNS - Dallas

4.5 THE OHIO STATE UNIVERSITY

The Ohio State University's Library Control System (LCS), one of the first operational online catalogs, is heavily utilized. We received from Ohio State log tapes for all of 1981, in addition to tapes for March, April, May, and November 1982. We decided to use the 1981 data for a number of reasons.

First, we had the data early. We were able to examine it immediately and determine methodology for reduction, parsing, etc. Second, as part of a terminal queuing study conducted for National Science Foundation, we had observers stationed at the terminals to record actual user arrival times. Thus, we were given the opportunity to determine user sessions by observation. Since none of the systems had time stamps to mark the beginning and end of sessions (LC was only partially successful in this regard as explained earlier), direct observation of LCS terminals provided us with the only reliable session data. Third, we had much more extensive data over time and could compare statistics from separate academic quarters.

Characteristics of the tape data were discussed in section 2. Basically the three types of outcome for a search of LCS are:

1. Subject index records
2. Author, title, date records
3. Display records

SIS and PS paging commands return subject index records, and AUT, ATS, TLS, SPS, SBL, and the PG paging commands return author, title, and date records. The display records are of two types: circulation records and bibliographic records. DSC, DST, and DSL return circulation records, while FBC, FBT, and FBL commands return bibliographic records. The PD paging command returns either circulation or bibliographic records, depending on the previous command.

Thus, in developing a map for the OSU data, we have three basic states that may be further subdivided. SIS makes up one state; the PS command is its associated paging command. The line select, SBL, allows one to go from the type 1 outcome to type 2. AUT, ATS, TLS, and SPS make up another state, and the PG command is its associated paging command. The line select commands, DSL and FBL, allows one to go from a type 2 to a type 3 display.

DSC, DST, FBC, and FBT may move a user up to yet another state, and the PD commands are their associated paging commands.

4.5.1 Transition Probability Matrices, State Diagrams, and Statistics

Tables 40 through 46 represent state transition matrices for two different mappings of the Ohio State data. Figures 35 through 66 show the state diagrams. The second mapping represents a breakdown of the SRH (search) command code into SRH AUT, SRH ATS, etc., as well as more detailed breakdowns of other commands. State diagrams showing the percent of transitions from an origin state to a destination state are shown in Figures 35 through 66. The terms compatible and fifth alternate used on the matrices and state diagrams simply refer to different mappings that we used for Ohio State. All data is from spring 1981.

Table 40. State Transition Matrix
Compatible - 1st Order - Successor - OSU
(Expressed in Percent)

	BGNS	BRWS	PS	SRCH	PG	DISP	PD	ERRS	ENDS
BGNS	0	21.4	0.2	48.7	0.6	9.8	0.1	19.3	0
BRWS	0	23.1	17.4	33.7	0.7	1.8	0	18.8	4.6
PS	0	16.1	37.0	33.0	0.8	2.1	0.2	7.4	3.5
SRCH	0	6.2	1.5	31.8	10.0	34.6	0	8.6	7.3
PG	0	2.6	1.0	7.2	53.6	25.0	0.2	6.3	4.2
DISP	0	3.8	0.9	17.3	6.3	45.7	4.1	6.4	15.5
PD	0	2.3	1.4	11.4	2.3	28.2	36.2	2.0	16.2
ERRS	0	15.7	2.8	26.1	5.5	5.0	0.8	33.3	10.8
ENDS	100.0	0	0	0	0	0	0	0	0

Table 41. State Transition Matrix
Compatible - 2nd Order - Successor - OSU
(Expressed in Percent)

	BGNS	BRWS	PS	SRCH	PG	DISP	PD	ERRS	ENDS
BGNS	0	7.1	3.8	29.2	5.3	28.7	1.3	16.0	8.6
BRWS	0	28.6	15.2	21.6	4.3	9.6	0	13.3	7.3
PS	0	16.7	24.2	26.4	5.5	12.3	0.2	8.9	5.8
SRCH	0	6.2	2.5	34.3	10.9	23.5	1.7	9.3	11.5
PG	0	3.6	1.7	12.3	45.3	21.3	0.9	7.6	7.3
DISP	0	3.6	1.7	17.6	7.3	49.3	3.9	6.6	10.2
PD	0	4.0	1.7	13.5	3.3	37.3	20.5	7.0	12.6
ERRS	0	10.6	4.6	22.3	7.1	15.9	0.6	30.0	8.8
ENDS	0	0	0	0	0	0	0	0	0

Table 42. State Transition Matrix
Compatible - 3rd Order - Successor - OSU
(Expressed in Percent)

	BGNS	BRWS	PS	SRCH	PG	DISP	PD	ERRS	ENDS
BGNS	0	8.6	4.5	23.9	7.1	24.7	3.1	13.9	14.2
BRWS	0	25.8	11.9	23.6	6.0	11.1	0.1	13.7	7.8
PS	0	17.4	20.6	24.6	9.2	13.5	0.3	9.4	5.1
SRCH	0	6.4	3.3	31.6	11.4	26.2	2.0	9.7	9.4
PG	0	3.9	2.0	14.2	41.0	22.1	1.2	8.5	7.1
DISP	0	3.9	2.3	19.5	8.5	46.7	2.7	6.9	9.6
PD	0	3.3	2.1	14.8	3.3	29.8	23.3	8.9	14.5
ERRS	0	10.9	5.3	21.5	8.6	16.9	1.0	25.1	10.7
ENDS	0	0	0	0	0	0	0	0	0

Table 43. State Transition Matrix
Compatible - 4th Order - Successor - OSU
(Expressed in Percent)

	BGNS	BRWS	PS	SRCH	PG	DISP	PD	ERRS	ENDS
BGNS	0	8.4	3.7	23.0	8.6	27.9	3.0	12.6	12.7
BRWS	0	25.1	11.1	22.9	7.8	12.0	0.4	13.5	7.4
PS	0	16.7	21.5	21.0	11.4	14.3	0.5	9.8	4.8
SRCH	0	6.8	3.5	32.7	11.6	24.9	1.8	9.3	9.4
PG	0	4.5	2.5	15.1	37.6	23.1	1.1	9.0	7.2
DISP	0	4.0	2.5	19.0	9.9	45.7	2.8	7.4	8.7
PD	0	3.9	3.3	15.7	5.3	33.0	18.3	6.5	14.1
ERRS	0	10.0	5.0	21.6	9.1	18.9	1.1	24.0	10.3
ENDS	0	0	0	0	0	0	0	0	0

Table 44. State Transition Matrix
Fifth Alternate - 1st Order - Successor - OSU
(Expressed in Percent)

	BEGIN	SRH SIS	BRW PS	SRH SBL	SRH AUT	SRH ATS	SRH TLS	SRH SPS	BRW PG	DSL/FBL	DSC/FBC	DIS PD	ERRORS	END S
BEGIN	0	21.4	0.2	0.3	13.6	7.7	26.3	0.7	0.6	2.0	7.7	0.1	19.3	0
SRH SIS	0	23.1	17.4	31.8	0.4	0	1.3	0.1	0.7	1.6	0.2	0	18.8	4.6
BRW/PS	0	16.1	37.0	30.7	0.5	0.2	1.6	0.1	0.8	1.9	0.2	0.2	7.4	3.5
SRH SBL	0	17.6	4.9	24.2	0.9	0.4	1.2	0.3	9.5	20.4	2.4	0	9.5	8.7
SRH AUT	0	1.6	0.1	0	25.8	2.6	7.7	0.1	13.9	30.3	0.3	0	10.3	7.2
SRH ATS	0	0.3	0	0	5.7	18.1	10.1	0	2.6	53.4	0.2	0	4.3	5.4
SRH TLS	0	1.8	0	0.1	4.3	2.0	23.1	0.1	9.8	42.9	0.6	0.1	8.7	6.8
SRH SPS	0	0.7	0.2	0.5	0	0.5	2.0	57.1	15.0	7.4	5.9	0	4.4	6.2
BRW/PG	0	2.6	1.0	0.9	2.1	0.8	2.7	0.7	53.6	24.4	0.5	0.2	6.3	4.2
DSL/FBL	0	3.8	1.0	2.2	4.8	3.6	7.7	0.3	7.2	43.0	1.7	3.0	6.5	15.3
DSC/FBC	0	4.1	0.4	1.1	1.4	1.7	3.4	2.7	1.4	3.2	48.0	10.6	5.4	16.6
DIS/PD	0	2.3	1.4	0.6	1.8	0.9	7.9	0.1	2.3	8.3	19.9	36.2	2.0	16.2
ERRORS	0	15.7	2.8	5.3	7.2	2.5	10.6	0.4	5.5	3.5	1.6	0.8	33.3	10.8
END S	100.0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 45. State Transition Matrix
Fifth Alternate - 2nd Order - Successor - OSU
(Expressed in Percent)

	BEGIN	SRH SIS	BRW PS	SRH SBL	SRH AUT	SRH ATS	SRH TLS	SRH SPS	BRW PG	DSL/FBL	DSC/FBC	DIS PD	ERRORS	END S
BEGIN	0	7.1	3.8	7.3	6.9	3.3	11.3	0.3	5.3	25.1	3.6	1.3	16.0	0
SRH SIS	0	28.6	15.2	18.4	0.8	0.4	1.8	0.3	4.3	8.3	1.3	0	13.3	7.3
BRW PS	0	16.7	24.2	23.6	0.7	0.1	2.1	0.1	5.5	11.0	1.3	0.2	8.9	5.8
SRH SBL	0	16.0	7.4	30.1	1.0	0.3	1.6	0.3	9.5	14.7	1.6	0.3	9.9	7.3
SRH AUT	0	2.2	0.6	0.5	23.1	2.6	7.8	0.4	15.2	24.4	0.7	0.3	10.3	11.8
SRH ATS	0	2.2	0.2	0.4	5.7	22.8	7.8	0.1	5.0	29.7	1.1	0.4	6.8	17.8
SRH TLS	0	2.1	0.5	0.6	5.3	2.1	23.0	0.2	11.0	26.0	1.5	4.3	9.6	13.8
SRH SPS	0	2.6	0	0.8	0.3	0.5	1.3	60.6	14.2	7.3	4.7	1.0	4.2	2.4
BRW PG	0	3.6	1.7	2.3	3.6	1.2	4.6	0.5	45.3	20.1	1.2	0.9	7.6	7.3
DSL/FBL	0	3.6	1.8	4.0	4.5	2.8	6.7	0.4	8.2	47.2	1.5	2.8	6.9	9.7
DSC/FBC	0	3.6	0.7	3.3	1.9	1.1	4.3	2.1	2.2	7.9	45.0	10.2	4.9	12.9
DIS PD	0	4.0	1.7	1.8	1.2	0.8	9.0	0.8	3.3	15.4	21.9	20.5	7.0	12.6
ERRORS	0	10.6	4.6	8.5	4.9	1.7	7.0	0.2	7.1	14.4	1.6	0.6	30.0	8.8
END S	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 46. State Transition Matrix
Fifth Alternate - 3rd Order - OSU
(Expressed in Percent)

	BEGIN	SRH SIS	BRW PS	SRH SBL	SRH AUT	SRH ATS	SRH TLS	SRH SPS	BRW PG	DSL/FBL	DSC/FBC	DIS PD	ERRORS	END S
BEGIN	0	8.6	4.5	5.1	5.7	2.9	9.8	0.3	7.1	21.3	3.4	3.1	13.9	14.2
SRH SIS	0	25.8	11.9	20.1	1.1	0.3	1.8	0.3	6.0	9.6	1.5	0.1	13.7	7.8
BRW PS	0	17.4	20.6	21.4	0.9	0	2.1	0.1	9.2	12.3	1.2	0.3	9.4	5.1
SRH SBL	0	16.3	8.7	27.9	0.8	0.4	1.8	0.4	9.3	14.5	1.8	0.5	10.8	6.7
SRH AUT	0	2.2	0.8	1.0	19.9	2.2	7.6	0.3	15.3	28.0	0.8	0.5	10.6	10.9
SRH ATS	0	2.1	0.8	1.2	5.7	17.1	8.2	0	7.4	38.2	0.8	1.5	6.7	10.4
SRH TLS	0	2.1	1.1	1.0	5.5	2.4	19.0	0.3	12.0	30.0	1.4	4.5	9.7	10.9
SRH SPS	0	1.9	0.8	1.9	0.8	0	2.2	58.6	11.6	5.1	5.4	1.1	4.8	5.9
BRW PG	0	3.9	2.0	3.4	4.2	1.3	4.4	0.8	41.0	20.8	1.3	1.2	8.5	7.1
DSL/FBL	0	3.9	2.5	4.4	4.6	3.6	7.8	0.4	9.3	43.3	1.8	1.9	7.1	9.4
DSC/FBC	0	3.9	0.9	2.8	1.4	1.0	4.8	2.1	3.8	9.5	46.4	7.1	5.4	10.9
DIS PD	0	3.3	2.1	2.8	1.9	0.7	8.9	0.5	3.3	16.4	13.4	23.3	8.9	14.5
ERRORS	0	10.9	5.3	8.2	4.8	1.7	6.5	0.3	8.6	14.6	2.1	1.0	25.1	10.7
END S	0	0	0	0	0	0	0	0	0	0	0	0	0	0

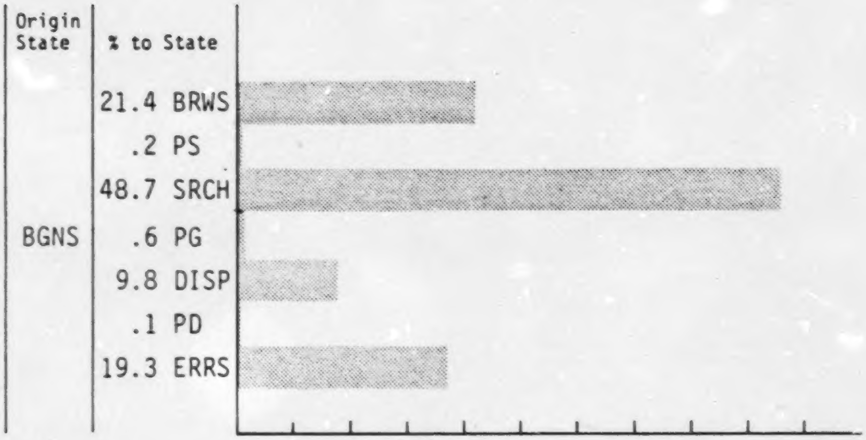


Figure 35. State Transition Diagram from BGNS State Compatible - 1st Order - OSU

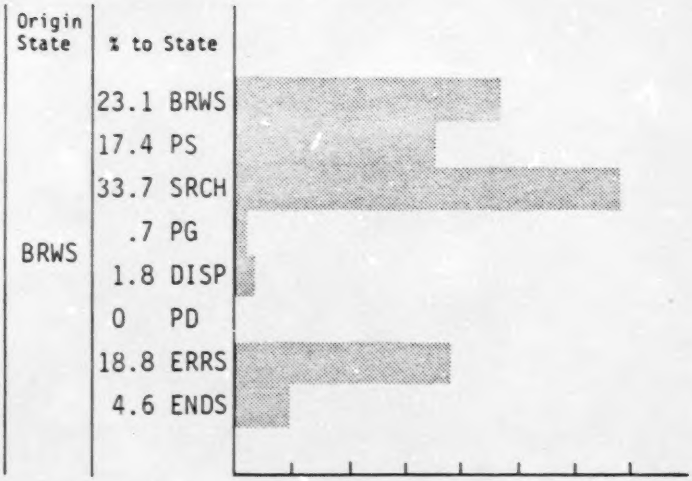


Figure 36. State Transition Diagram from BRWS State Compatible - 1st Order - OSU

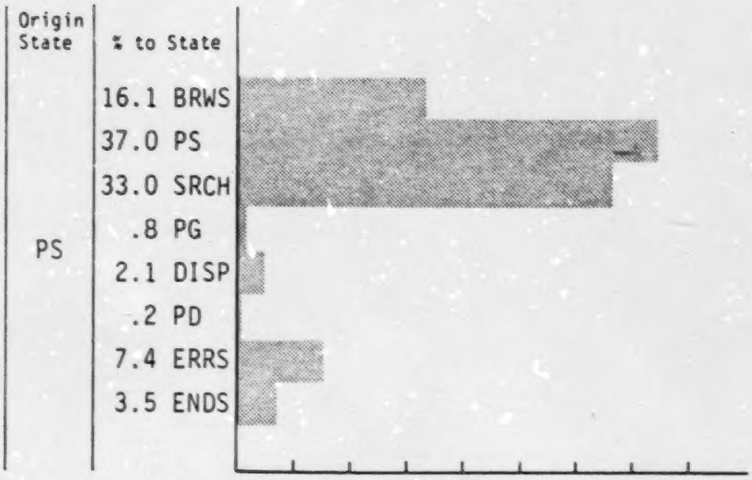


Figure 37. State Transition Diagram from PS State Compatible - 1st Order - OSU

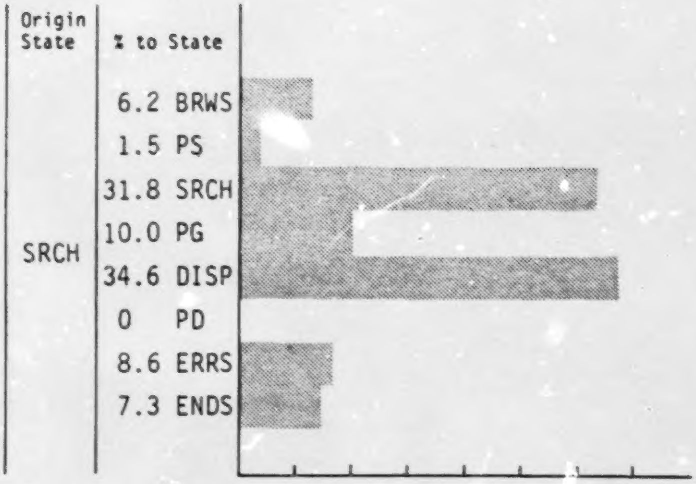


Figure 38. State Transition Diagram from SRCH State Compatible - 1st Order - OSU

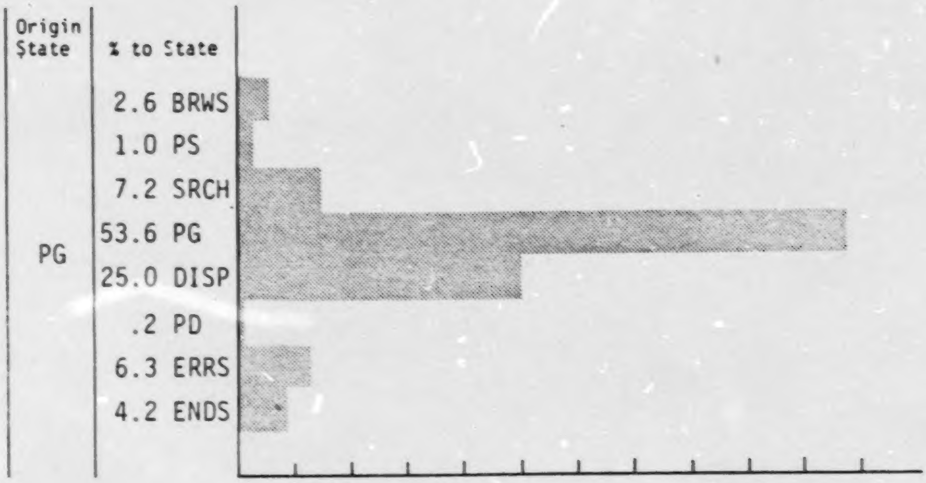


Figure 39. State Transition Diagram from PG State
Compatible - 1st Order - OSU

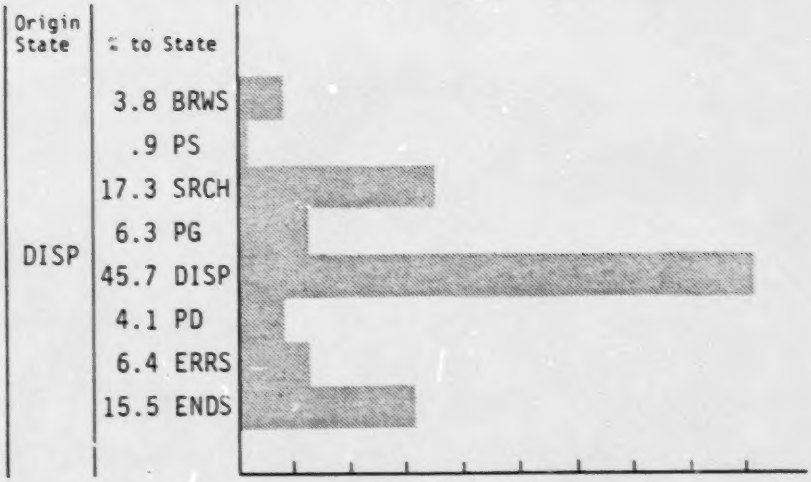


Figure 40. State Transition Diagram from DISP State
Compatible - 1st Order - OSU

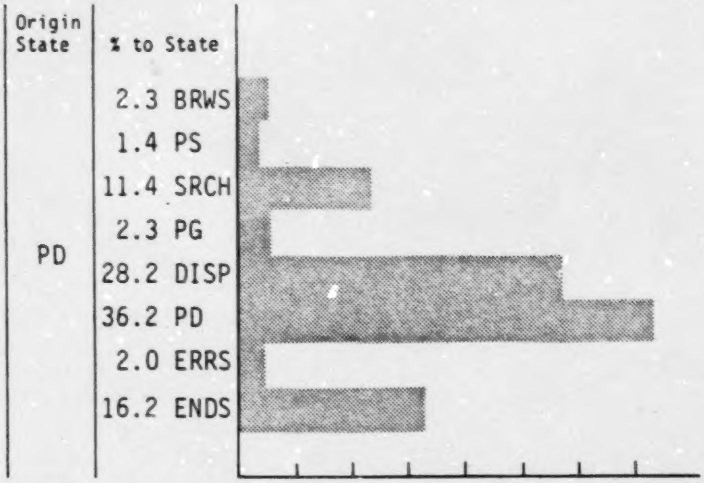


Figure 41. State Transition Diagram from PD State
Compatible - 1st Order - OSU

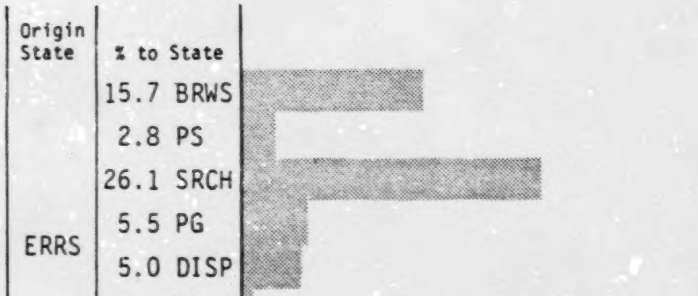


Figure 42. State Transition Diagram from ERRS State
Compatible - 1st Order - OSU

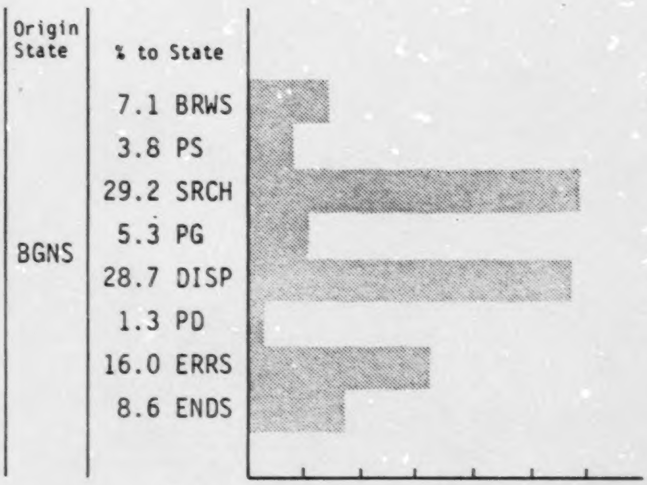


Figure 43. State Transition Diagram from BGNS State
Compatible - 2nd Order - OSU

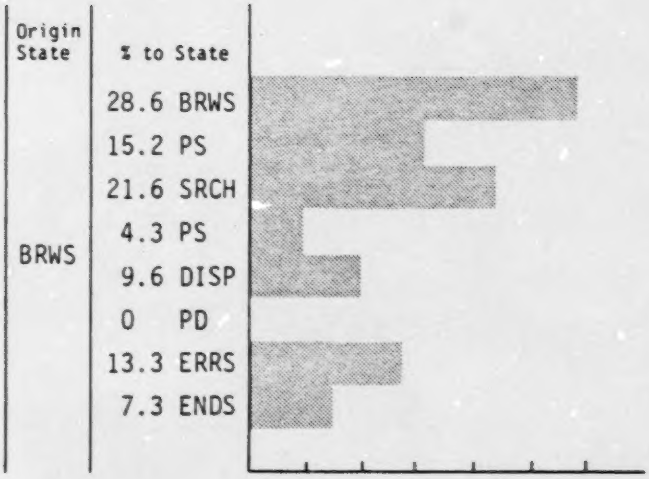


Figure 44. State Transition Diagram from BRWS State
Compatible - 2nd Order - OSU

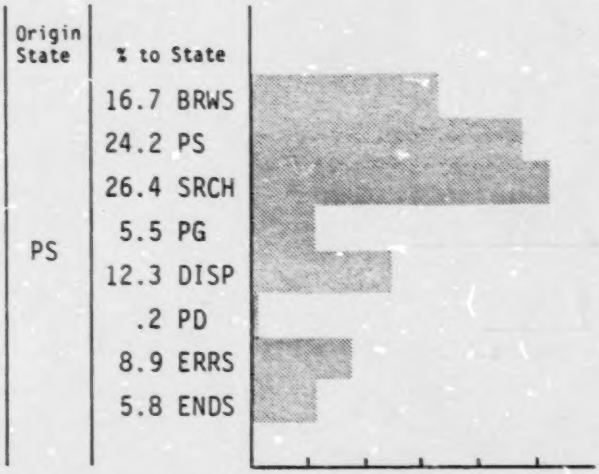


Figure 45. State Transition Diagram from PS State
Compatible - 2nd Order - OSU

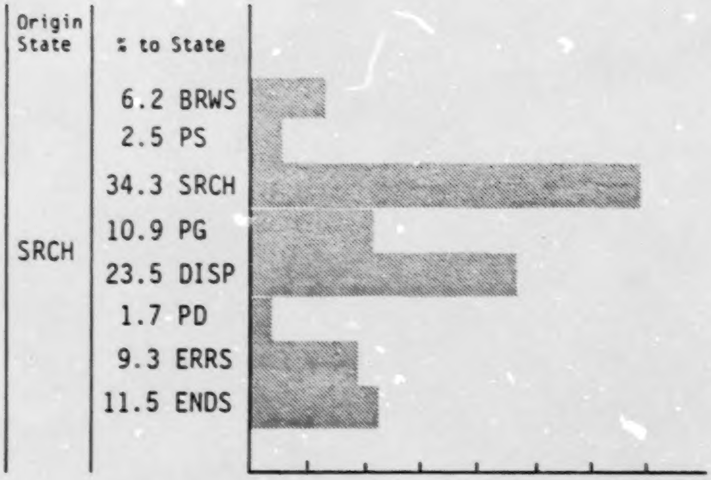


Figure 46. State Transition Diagram from SRCH State
Compatible - 2nd Order - OSU

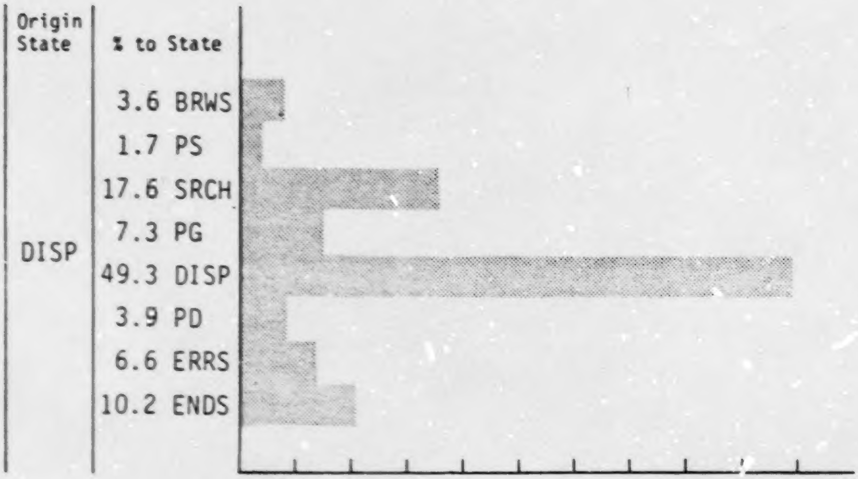


Figure 47. State Transition Diagram from PG State
Compatible - 2nd Order - OSU

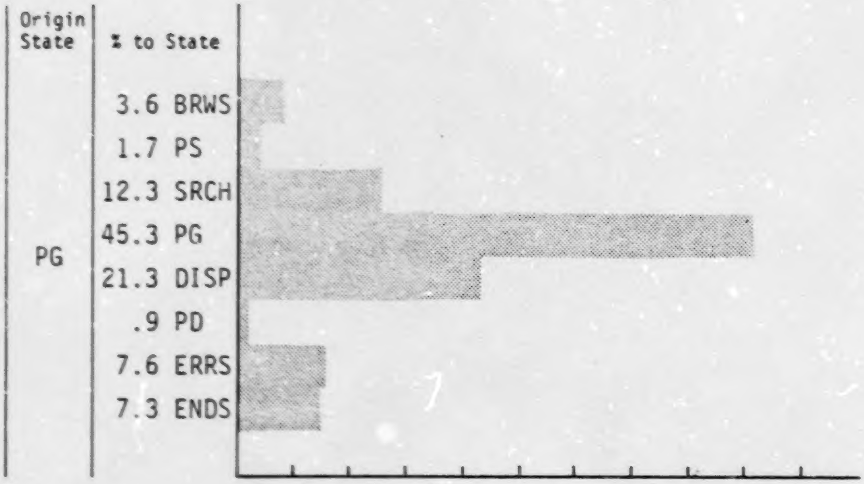


Figure 48. State Transition Diagram from DISP State
Compatible - 2nd Order - OSU

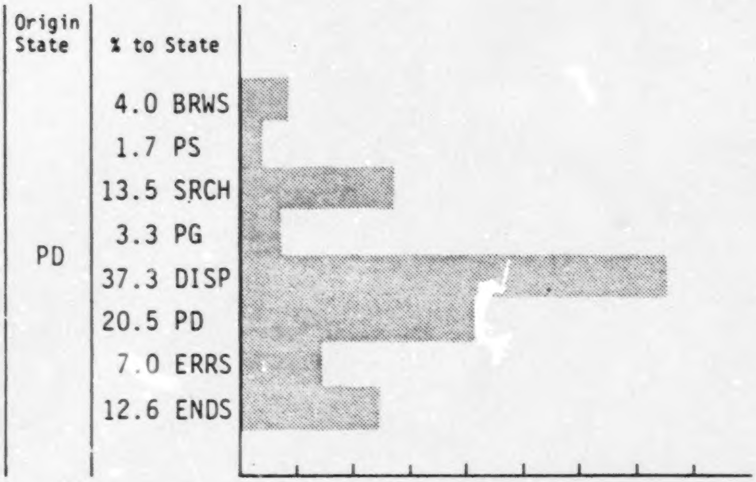


Figure 49. State Transition Diagram from PD State
Compatible - 2nd Order - OSU

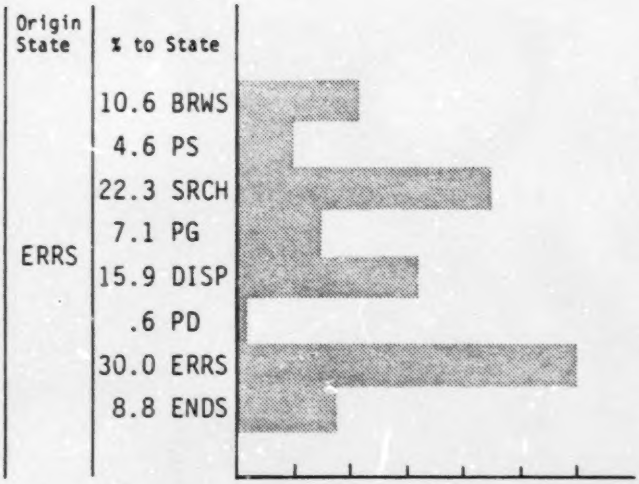


Figure 50. State Transition Diagram from ERRS State
Compatible - 2nd Order - OSU

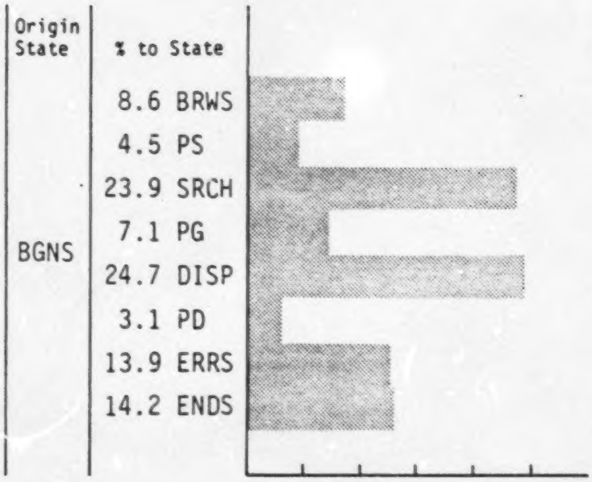


Figure 51. State Transition Diagram
3rd Order from BGNS - OSU

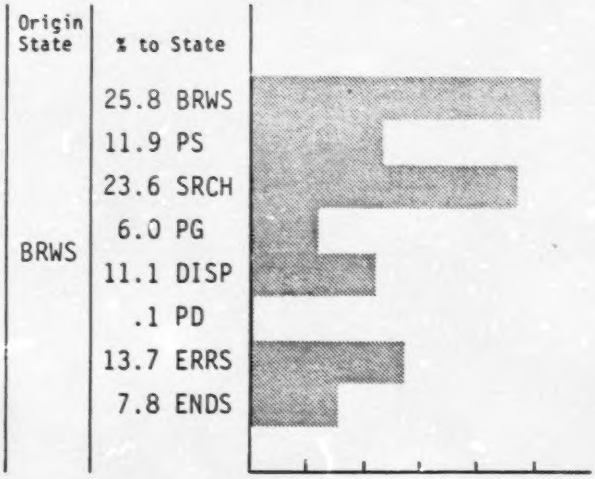


Figure 52. State Transition Diagram
3rd Order from BRWS - OSU

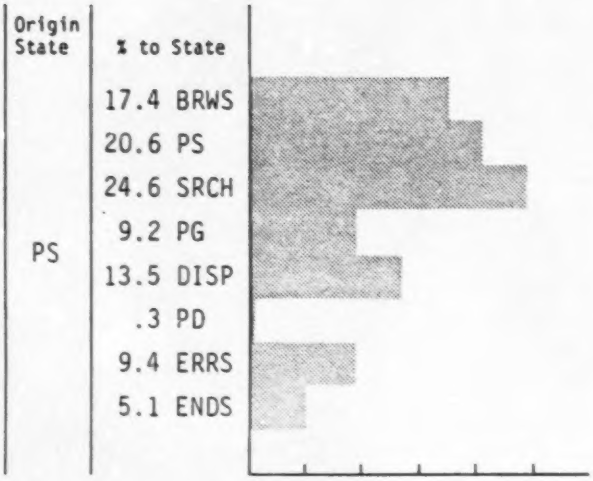


Figure 53. State Transition Diagram
3rd Order from PS - OSU

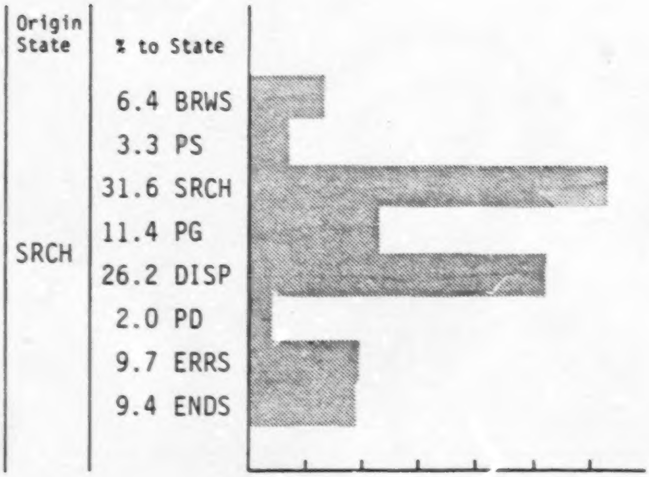


Figure 54. State Transition Diagram
3rd Order from SRCH - OSU

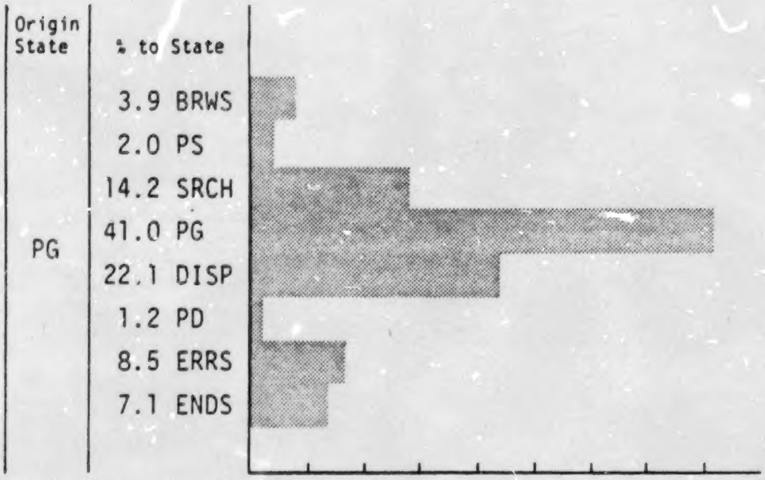


Figure 55. State Transition Diagram
3rd Order from PG - OSU

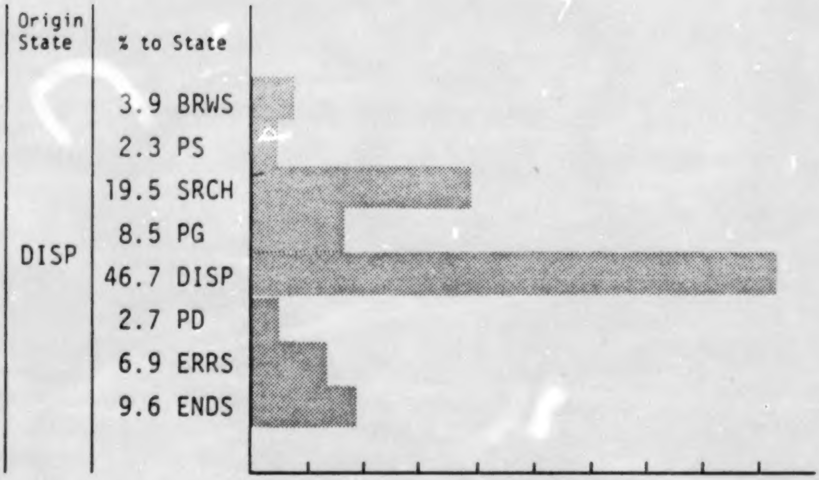


Figure 56. State Transition Diagram
3rd Order from DISP - OSU

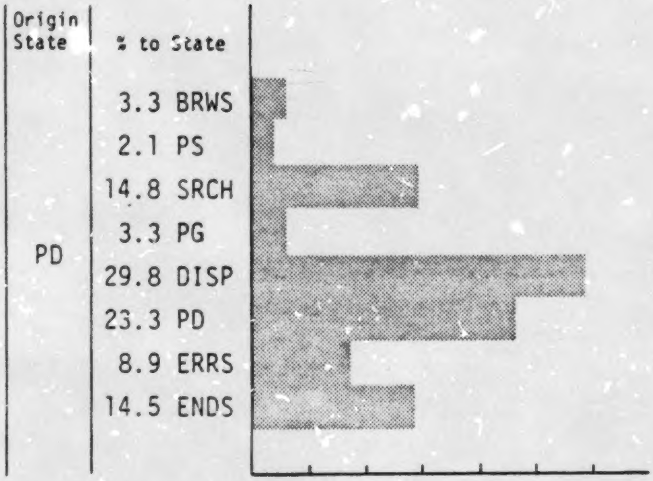


Figure 57. State Transition Diagram
3rd Order from PD - OSU

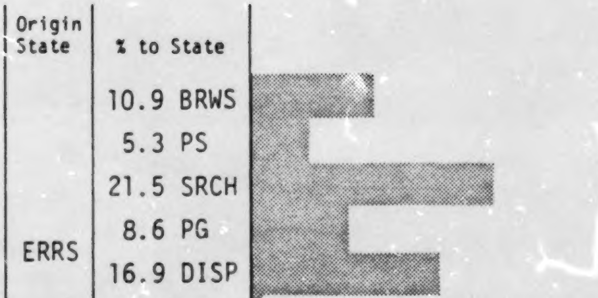


Figure 58. State Transition Diagram
3rd Order from ERRS - OSU

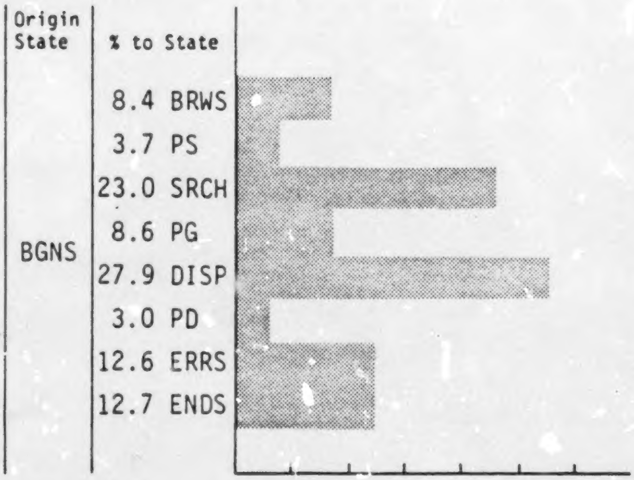


Figure 59. State Transition Diagram
4th Order from BGNS - OSU

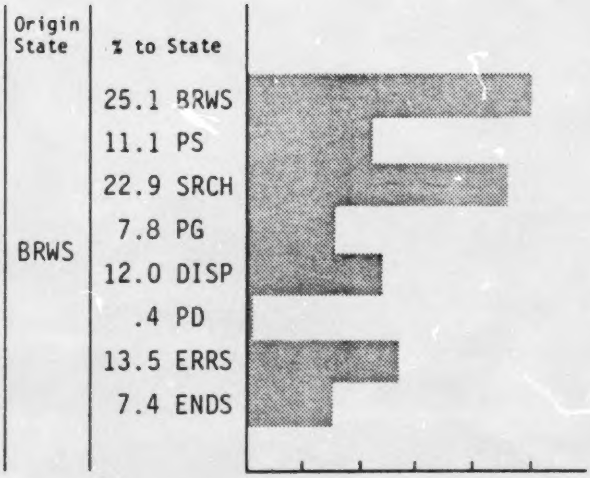


Figure 60. State Transition Diagram
4th Order from BRWS - OSU

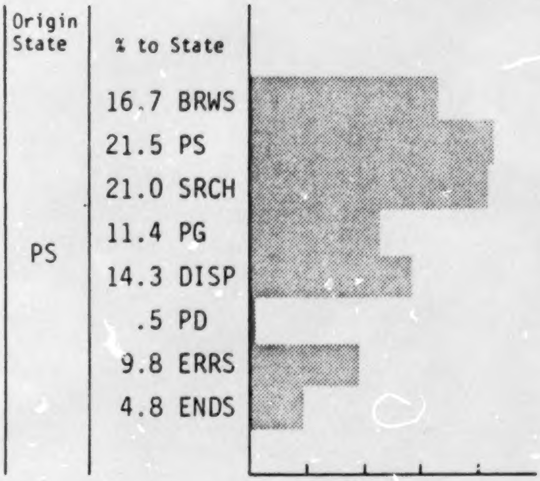


Figure 61. State Transition Diagram
4th Order from PS - OSU

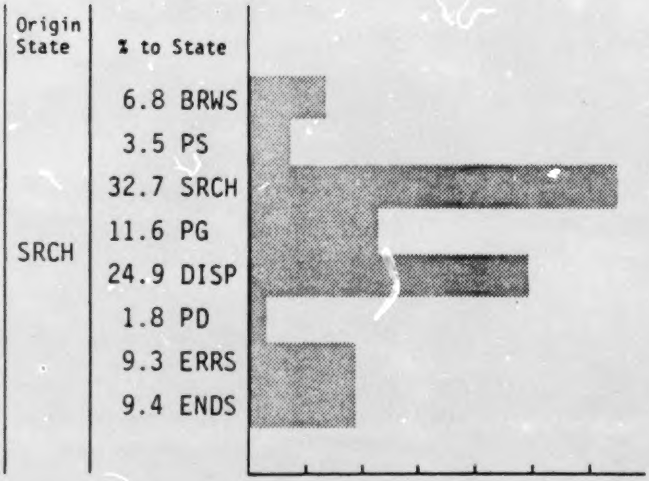


Figure 62. State Transition Diagram
4th Order from SRCH - OSU

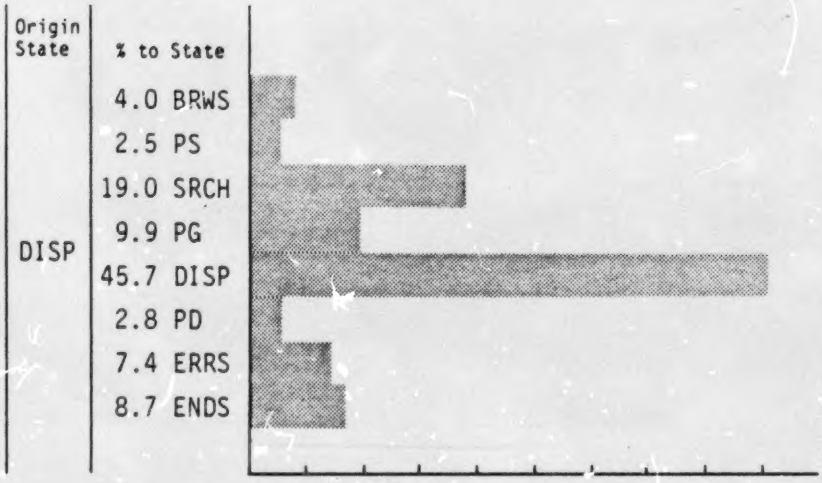


Figure 63. State Transition Diagram
4th Order from PG - OSU

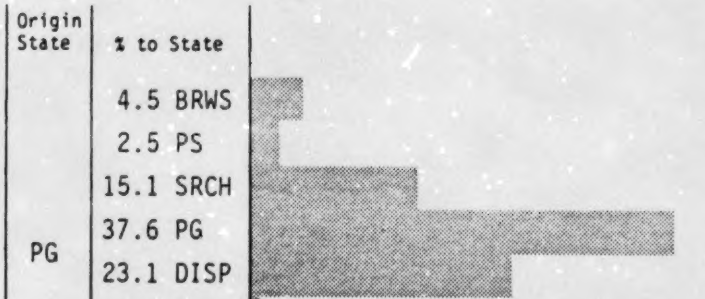


Figure 64. State Transition Diagram
4th Order from DISP - OSU

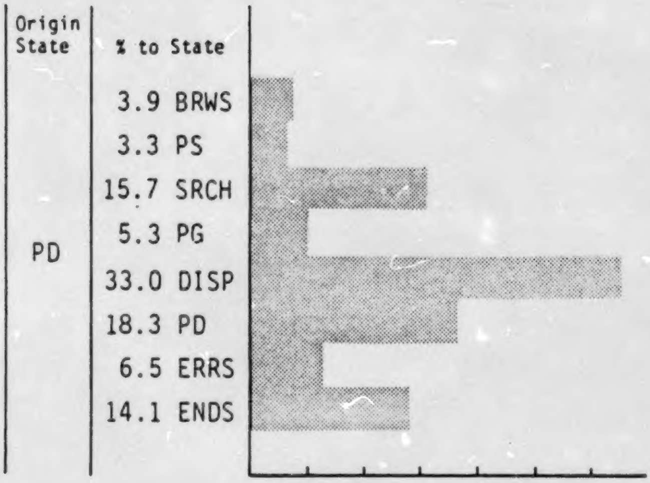


Figure 65. State Transition Diagram
4th Order from PD - OSU

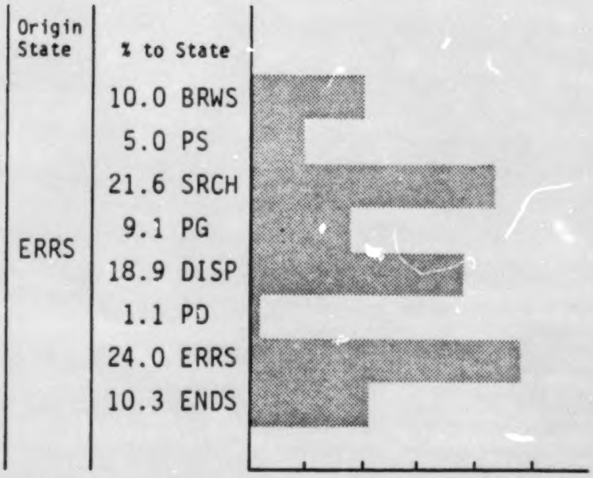


Figure 66. State Transition Diagram
4th Order from ERRS - OSU

4.6 OCLC PUBLIC TERMINALS

Although access to bibliographic records on OCLC terminals was not designed for the untrained user, users express highly favorable attitudes toward the OCLC Online Union Catalog compared with attitudes toward online catalogs designed expressly for library patrons. A higher percentage of OCLC users responded with "Very favorable" or "Somewhat favorable" than did users of other online catalogs when asked this question on the CLR Survey (discussed by Markey in Volume II, section 5: "My general attitude toward the computer catalog is:".)

All transactions on OCLC public terminals are captured and recorded daily, regardless of the subsystem to which they are logged-on. For this study, all public terminal transactions from May through October 1982 were extracted from the daily transaction logs and gathered together for analysis. The data categories identified for this analysis include nine bibliographic search types. The number of public terminals used on a given day and all transactions that were not bibliographic database searches were grouped into the "other" category.

The number of OCLC public terminals used varies from day-to-day for a number of reasons. For example, one or more libraries in this group may be closed, or a specific terminal may not have been used. The average number of public terminals in daily use throughout this six-month period is shown in Figure 67. Since 81% of the libraries providing OCLC terminals for public searching are academic libraries, the September increase of terminals in use is not surprising. However, the June increase stimulates reexamination of our assumption that certain kinds of library activity diminish when the summer season arrives.

For the entire six-month period, the ratio of bibliographic searches to all other transactions conducted at these OCLC public terminals was 2.1 (other) to 1 (searches). Figure 68 shows the average number of bibliographic searches and the average number of other transactions, per terminal per day, throughout the six-month period. Significant changes in this ratio occur in July and August. Less staff activity is suggested by the July data; more staff activity seems to have occurred in August.

Nine search types were included in the analysis of bibliographic searching on the public terminals: author/title, title, author, OCLC number, LCCN, ISBN, ISSN, CODEN, and government document number. Corporate author searches are included with author searches. Figure 69 shows the frequency (daily maximum, minimum, and average) of these nine search types for the entire six-month period. The relatively high frequency of OCLC number searches can be explained in part by the encouragement given to patrons to search for serial titles located on a printed list by OCLC number. The predominance of author and title searches supports the well-understood notion that both staff and patrons use OCLC for known-item searching.

Figures 70 and 71 should be viewed together, as they reflect the increases and decreases in each of the nine search types from May through October 1982. The changing rate of the average search transactions per day does not remain consistent across all nine search types. Author/title and title searches increase slightly in August, while OCLC number, LCCN, and ISSN searches decrease, on the average. LCCN searches were the only searches to decrease, on average, during September, the busiest month of this period.

Most users of online catalogs have used only a single system. Existing online catalogs vary considerably in design philosophy and access points. Each of the libraries providing patrons with OCLC public terminals also maintains one or more additional forms of catalog access to its collections. For these reasons, it is too early to make comparisons among different forms of catalog access or evaluate one mode of access solely on the basis of activity reports or users' stated preferences. The data summarized here should be viewed merely as an indication of how OCLC's public, multipurpose terminals were used during the period (May through October 1982) in which data was collected.

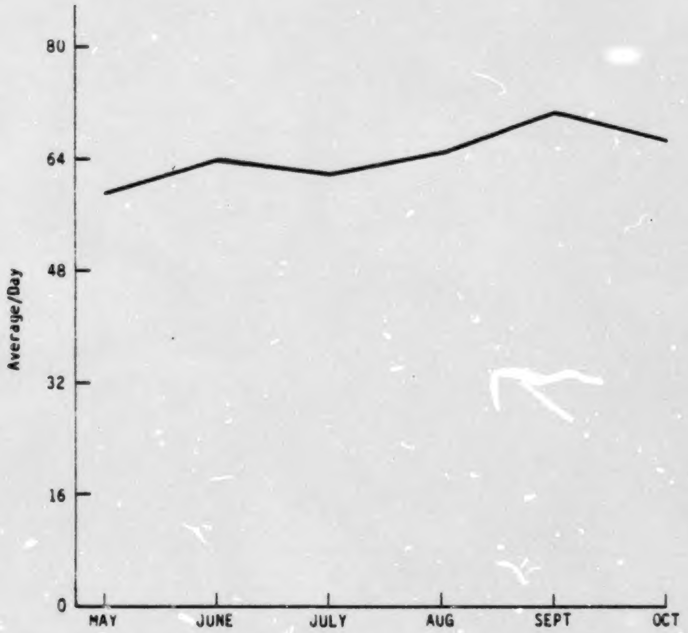


Figure 67. OCLC Public Terminals in Use

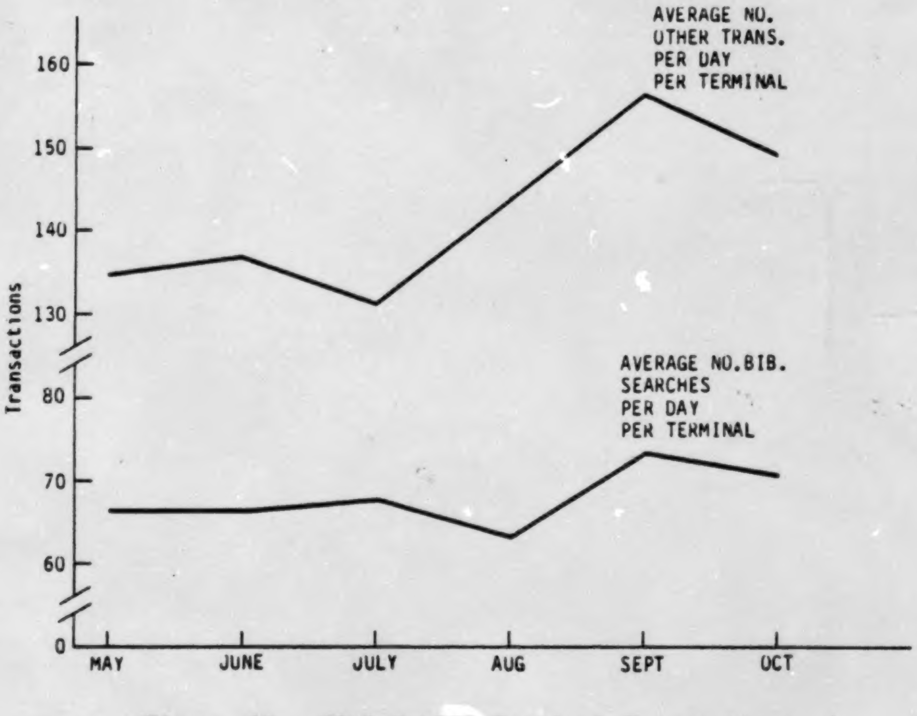


Figure 68. OCLC Public Terminal Transactions

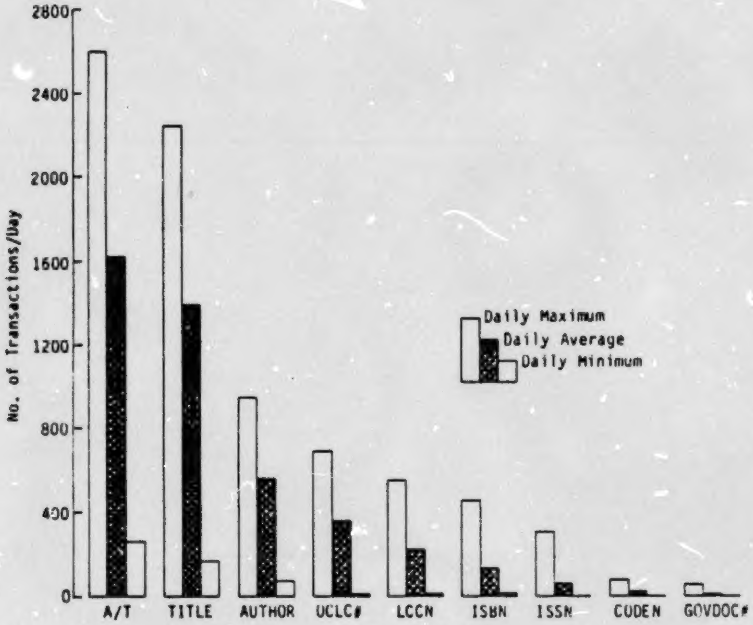


Figure 69. Frequency of OCLC Searches by Type
(May-October 1982)

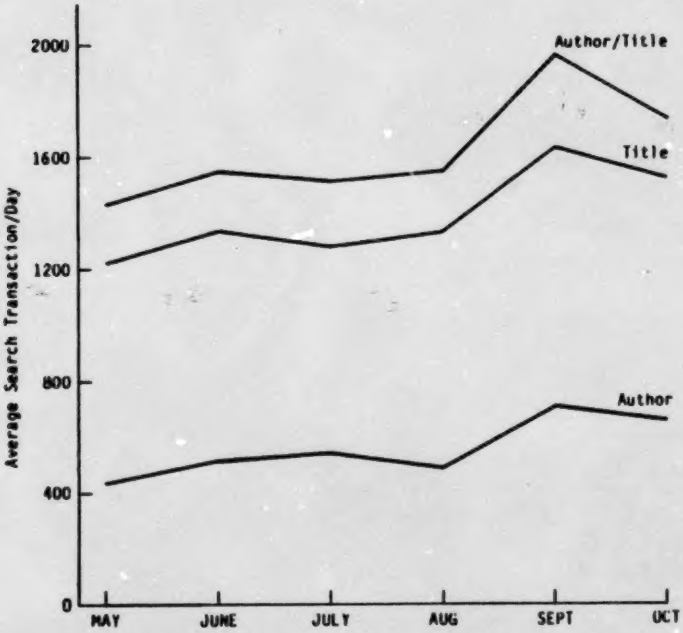


Figure 70. OCLC Public Terminals: Author and Title Searches
(May-October 1982)

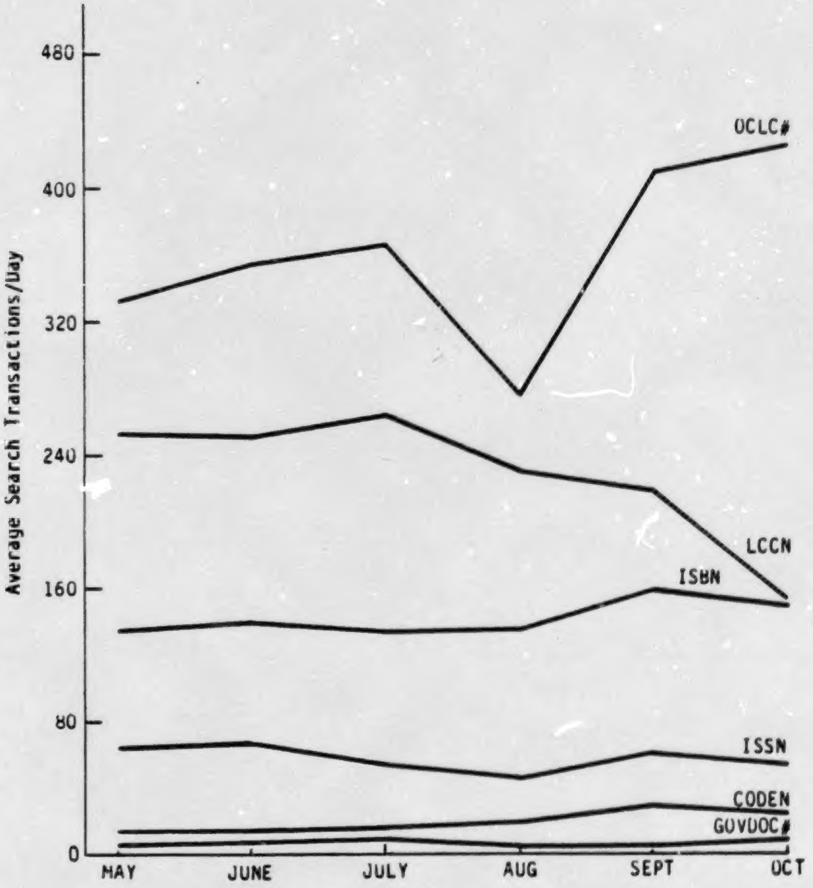


Figure 71. OCLC Public Terminals: Standard Number Searches (May-October 1982)

5.0 CONCLUSIONS, OBSERVATIONS, AND RECOMMENDATIONS FOR FURTHER STUDY

A number of interesting and useful observations can be drawn from the observed statistics. The questions suggested in section 3.1 are addressed below. In addition to these questions, a noninclusive listing of observations and conclusions is presented that may provide further insight into what is really happening at online catalogs.

5.1 QUESTIONS AND ANSWERS

QUESTION 1. How much time did patrons spend at the terminal?

ANSWER For the Library of Congress, we do have the number of commands per session. The median session has between 9 and 10 commands, but no time stamps are available, thus we do not know anything about elasped time within a session. The longest session was 292 commands, and the shortest was 2 commands (by definition, a begin and end command). Approximately one-third of the sessions have 6 or fewer commands, one-third between 7 and 16 commands, and one-third 17 or more commands.

For OSU, the average terminal session length ranged from just under 4 minutes to over 9 minutes, depending on the library branch. The median time that patrons took to formulate the following commands was:

Author/title	34 seconds
Author	31 seconds
DSC	30 seconds
DSL	16 seconds
FBL	17 seconds
SIS (subject)	28 seconds
TLS	31 seconds

For Syracuse, the median session time was 3.55 minutes, and the mean session time was 6.1 minutes. The median time that patrons took to formulate the following commands was:

Author to Author	16 seconds
Title to Title	20 seconds
Subject to Subject	19 seconds
Keyword to Keyword	19 seconds
Dual to Dual	27 seconds

For Dallas, the time was recorded only to the minute, but the median session was around 3 minutes.

QUESTION 2. What commands did patrons use?

ANSWER The use commands are reflected in the zero order statistics and higher order transitions. A general statement cannot be made since systems have different search commands and structures. However, an examination of the results for individual systems will give these statistics. These are found in section 4 for each OPAC system.

QUESTION 3. How did time per transaction vary by command?

ANSWER For OSU, the median time per transaction when a hit occurred was as follows:

ATS to ATS	71 seconds
ATS to AUT	70 seconds
ATS to SIS	103 seconds
ATS to TLS	87 seconds
ATS to DSC	132 seconds

Similar time patterns occurred for AUT to other commands. The time from DSC to others was considerably lower, from 22 to 38 seconds. From SIS to other commands, the range was 54 (for SIS), and from 84 to 123 for others. TLS to others ranged from 65 to 128 seconds. This is also addressed in question 2 for OSU. Such data was not available for Dallas or LC and was not determined for Syracuse.

QUESTION 4. What errors did people make? Where in the session did they make them?

ANSWER There is a definite tendency to repeat errors, i.e., once an error is made, there is a high probability of transition to another error. For Dallas, error occurrence was low, perhaps because of its format-filled layout and a corresponding tendency to eliminate syntactical errors. At LC, we had no indication of system response, and errors were defined as search codes that were not recognized. The most prevalent of these were the entry of a search key without a search code and miskeyed codes. For OSU, we know where in the session the error occurred, but we don't know what type of error. We do know, however, that single transaction sessions had a high error rate (41% of the sessions). It seems that errors have a slight tendency to decrease as session length increases.

QUESTION 5. How does behavior vary according to session length?

ANSWER For OSU, there seem to be fewer errors made as session lenght increases. It would be possible to examine the LC data to determine any basic frequency changes, but it was not done for this study. At LC, we segregated the data into short, medium, and long sessions. Short sessions contained from 2 to 6 commands; medium, from 7 to 16; and long, from 17 to 282. As expected, the beginning and ending commands decreased in proportion to overall commands as session length increased. The proportion of BRWS state commands decreased slightly while the proportion of FIND increased. The proportion of SRCH state commands rose as we went from short- to medium-length sessions and remained fairly constant as we went to long sessions. The most dramatic change was with the DISP state which rose as session length increased. In contrast with OSU, the error state remained constant across the different session lengths. However, it must be remembered that "error" at LC, since we do not have a system response, denotes a command we could not recognize, while at OSU an "error" was an error as recognized by the system. Hence, syntax errors, which are only a subset of all errors, remained constant at LC across the different lengths of session.

QUESTION 6. How does use vary among libraries (within the same institution)?

ANSWER At OSU, we observed different behavior with respect to session lengths at different libraries. Syracuse and Dallas were not subdivided by location, and LC represented the same location.

QUESTION 7. How does use vary over time (such as academic quarters)? Within libraries (in the same system)?

ANSWER Data for OSU covers a whole year, but only the spring quarter was analyzed. We did not have data for the other systems over a longer period of time.

QUESTION 8. What is the frequency of "hits"/no "hits"?

ANSWER Hit frequency was not available for LC. For Syracuse, Table 34 shows the percentage of 0, 1, 2, 3-4, 5-8, 9-16, 17-32, etc., hits per type of search. It illustrates that subject search is the least successful, since 52.8% of the searches resulted in zero hits. Author searches seem to result in the most multiple hits.

QUESTION 9. Given that the searcher is in a search type, such as author search, what is the likelihood of changing search type?

ANSWER For Syracuse, Table 29 illustrates the likelihood of going to different search types, given that the user is in a certain search type.

For Dallas, the percentages of switch to other searches from one search were:

SUBJ to other type 15%;
AUTH to other type 45%;
TITL to other type 48%;
CALL # to other type 41%;
AUTL to other type search 51%.

QUESTION 10. Are there sophisticated searching features available that are used infrequently? What correlation, if any, does type of library have with needed prompting devices or search sophistication?

ANSWER Novice and sophisticated users exist at all libraries. At LC, the search features of NEWS, SET, etc., are used infrequently. FIND is used frequently, but SCAN and RTR are not. If help is considered a sophisticated feature, then the HELP command is used less than 0.05% of the time at Syracuse. The Dual-Boolean is used at Syracuse about 5% of the time. However, many of these searches may be a Boolean author search instead of a title search, since there is no AUT/TITLE search available.

QUESTION 11. What is the relative frequency of available fields searched?

ANSWER The relative frequency of available fields searched can be observed from the zero order statistics.

QUESTION 12. What is the relationship between rigid command syntax, command usage, and error message frequency?

ANSWER Our data does not give this information (unless the Dallas menu-driven system is compared, in which case the Dallas system has few errors).

QUESTION 13. What, if any, patterns of feature use exist?

ANSWER The patterns of feature use can be observed in the transition matrices and state diagrams.

QUESTION 14. How does user time at the terminal compare with each type of search (e.g., author vs. subject)?

ANSWER This was not available for Syracuse, LC, or Dallas.

QUESTION 15. If an error is made, what is the likelihood of continuing to make an error?

ANSWER This is addressed in question 4.

QUESTION 16. Is there any relationship between errors and length of search?

ANSWER This was addressed in questions 4 and 5.

QUESTION 17. What are the frequencies of single-type search sessions and multiple-type search sessions?

ANSWER For OSU: 28.1% of searches were of one command only. 35.7% of searches were greater than one search type, and 36.2% were mixed search sessions. That is, two-thirds of the sessions were single-type searches. For Dallas: 65% were only one-type search, and 89% were one or two searches. For Syracuse, 43.6% were single searches.

5.2 OTHER OBSERVATIONS AND CONCLUSIONS

(1) For LC, Syracuse, and OSU, the 1st and 2nd order transitions between identical states have high probability (i.e., usually 1st or 2nd highest probability in the transition matrices). This implies that users tend to stay or return to the same state up to a state chain of 3 commands. In addition, analysis of 2nd orders tends to confirm that they remain in the same state throughout the transition chain. In particular, for OSU, the shelf position search (SPS) tends to remain through the 4th order (SPS is where users put in the call number and browse around it). Remaining in a given state may account for the majority of the higher order chains when the beginning and ending state of the chain is the same.

(2) For specific states, we observe that the transition from Display state to Display state is not necessarily important at SULIRS, since often the DISP command is used as a paging command through returns. An exception to remaining in Display is at Dallas, where the probability of remaining in DISP is low. When considerable information is shown on the screen (15 hits) and a detailed line is selected, then either more detail can be chosen, a browse done, or a new search begun. This occurs at LC, since it is the only way to look at individual records (paging).

All search states have a tendency to go to the same search (AUT, ATL, TITL, SUBJ, CALL, KEYW, DUAL, SERIES TITLE, BOOLEAN, etc.), and users stay in the same search for a long time.

(3) Subject searches are highest in 2 out of 3 systems that have them, but note that at Syracuse, which has subject search as highest (26% vs. AUT 24%), the 1st order transitions from the Begins State are:

Begins to AUT	25.1%
Begins to TITL	23.5%
Begins to SUBJ	21.9%
Begins to KWORD	17.6%

That is, more sessions begin with an author or title search than with a subject search. The "enhanced" number of subject searches may be a reflection of users who use such searches frequently within a session and not an indication of the "type" of search chosen. Another observation is that subject search fails more often than any other search, except the DUAL search, in terms of getting zero hits (i.e., it is not a particularly successful type of search). However, keyword is available at Syracuse, and should probably be included as a form of subject search. If we add subject and keyword, we then have 39.5%.

At The Ohio State University, we have the following transitions:

Begins to TITL	26.3%
Begins to SUBJ	21.4%
Begins to AUT	13.6%

However, this data has the error state contained in the analysis, and the figures are not directly comparable with those at Syracuse.

At Dallas, transitions from the beginning state are:

Begins to SUBJ	56%
Begins to TITL	19%
Begins to AUT	15%
Begins to CALL	5%
Begins to AUTL	4%

(4) At Syracuse, when repeated searches are factored out (for example, AUTH-AUTH, TITL-TITL, etc.), then AUTH-TITL is used twice as often as any other search type. Likewise, Title-Author is used more often. This indicates a close relationship between author and title. On the other hand, subject goes to Title, Author, and Keyword equally often, and Keyword goes to: SUBJ most often (23.5%); AUTH next (18.5%); and TITL third most often (14.6%).

(5) There is a trade-off between receiving transaction logs partially coded or completely "raw." Coding is useful up to the point of not losing vital information. For example, at Ohio State the terminal output was reflected in a one-byte return code, while at SULIRS no coding was done. When prior coding of command usage is not done, it is a major effort to code the actual data for analysis.

(6) Time stamps accurate to the second are vital for analyzing the characteristics of the system. Dallas had time stamps accurate only to

the minute, LC had neither time stamps nor date stamps, and both Syracuse and OSU had time stamps accurate to one-hundredth of a second and accurate date stamps.

(7) Since no codes were available from SULIRS, major efforts were necessary to code these. It is estimated that a minimum of 50% of the analysis time went into reducing and coding the data for subsequent analysis.

(8) Since no codes or system responses were available at LC, major efforts were also necessary here to reduce the data and code it meaningfully.

(9) The power of the analysis is directly dependent on the level of the captured information. We could conduct more in-depth analysis if we received more detailed data.

(10) When possible, it seems that it would pay to put effort into "upfront" data preparation, rather than attempting to extract the required information at the analysis stage. There are at least two reasons for this: first, we would be assured of obtaining the required information; and second, the level of data reduction and analysis effort would be greatly reduced over the long run.

(11) The problem of session determination is significant. No sessions were defined at OSU, Dallas, or Syracuse within the transaction log tapes. Two separate methods were used in an attempt to determine session boundaries. At Ohio State, data collectors were hired to observe users as they began their sessions at the terminals and to compare observations with log tapes. This method still required estimates for session endings and some session beginnings.

At Syracuse, we requested that people type S-blank when they began a session. Although the majority of users did not use S-blank, of those who did, a significant proportion used it within the same session. Based on this experience, asking the patron to indicate the session start is not reliable. At Dallas, some session starts were apparently available to be captured, but we did not receive any boundaries in the transaction tapes. The boundaries were drawn by considering time gaps between commands (e.g., when 3 minutes were exceeded). In addition, it was possible to determine with some accuracy where false boundaries were inserted by looking at the data, but it was not possible to spot where boundaries were not inserted.

We have a log on and a log off in the LC SCORPIO data. Users must log on to one of the available files. To successfully go to another file, they must log off. It appears some would leave their terminal and clear the screen without logging off, while others would not. Session boundaries were determined by end session commands, but there are numerous factors that do not always allow accurate session determination. The session boundaries are not reliable, and this is compounded by the absence of time stamps.

(12) Generalization is difficult, since we did not have a random sample of data. It does not mean the data is not generalizable. However, it does mean we cannot claim that it is.

(13) The frequency of subject searches when including multiple search sessions is less than author or title. This indicates less use of subject searches than author or title in multiple search sessions. Perhaps author and title are a more "refined" type of search than subject and, thus, are used more in multiple search sessions. A single search session is most likely when there is a special number (LC #, etc.). On the other hand, series title and dual (Boolean) are least likely to stand alone.

5.3 TRANSACTION ANALYSIS GOALS

A. OBJECTIVE AND GOALS

Volume I of this study is concerned with the analysis of the transaction log records of online catalogs. The objective of this part of the online public access catalog study was to discover the extent to which current online catalog features are used and patterns of that use. While examining the actual catalog use, at least two goals became obvious: to obtain a better understanding of the use of online catalogs and to refine the methodologies used in carrying out any transaction analysis study.

The first goal, to obtain a better understanding of the actual online catalog use, may seem to be too large or vague, as the objective is contained within it. However, it is important to keep in mind that a valuable asset of the research is that it generates questions that were not previously considered. As such, it contributes to an understanding of the catalog use. This study has achieved this in the sense that considerable understanding of catalog use has been obtained. There is of course more to do and investigate, but we are much closer to understanding how the online catalog is used.

The second goal, to refine the methodologies, was certainly achieved. The use of transaction analysis--via development of transition probability matrices, state definitions, state diagrams, and Markov analysis--to create chains of user patterns certainly offers promise. This transaction analysis methodology is still in the early stages of refinement, however. Further refinement will occur as more systems are examined and more data is analyzed.

The scope of the transaction analysis included collecting data from four different online catalog systems. Machine-readable log tapes, which contained the user transactions in various formats and codes, were received from the following institutions: Library of Congress, Syracuse University, Dallas Public Library, and The Ohio State University. In addition, data was collected on the searching of 76 OCLC public terminals at 53 different libraries (as defined by unique OCLC location sysmbols). By institution type, these 53 libraries included 5 public libraries, 5 federal or state libraries and 43 academic libraries (including several medical, health, and law libraries).

5.4 FUTURE EFFORTS

There are a number of future efforts that require consideration. One such effort would be to determine what types of comparability factors would be appropriate and meaningful for different systems. The detailed exploration of a particular system, with a priori data information requirements, also offers much potential. Controlled experiments with systems would also be useful to determine the effectiveness and sensitivity of the system when changed. The variance of system use over time (long periods, such as one year) with well defined data also offers promise if user population is stable.

Future efforts could also include alternative modeling efforts that consist of different analyses. Such potential efforts include, but are not limited to, the following:

1. Markov analysis
2. Multidimensional scaling analysis
3. Origin-destination models
    a. trip assignment, e.g., gravity model, interviewing opportunities model, etc.
    b. modal split
4. Clustering algorithms (to possibly identify user types)
5. Simulation
6. Controlled experiments with terminal users